Bayesian Analysis of Multi-Stratum Designs and Probability-based Optimal Designs with Separation

Mohammad Lutfor Rahman

School of Mathematical Sciences

Queen Mary, University of London

A thesis submitted for the degree of

Doctor of Philosophy

2015

I would like to dedicate this thesis to my wife Lisa and lovely daughter Zahra.

Acknowledgements

> "To Allah belongs whatsoever is in the heavens and the earth. Verily, Allah, He is *Al-Ghani* (Rich, Free of all wants), Worthy of all praise. And if all the trees on the earth were pens and the sea (were ink wherewith to write), with seven seas behind it to add to its (supply), yet the (praiseworthy) Words of Allah would not be exhausted. Verily, Allah is All-Mighty, All-Wise." Surah Luqman, Al-Quran [31:26-27].

My all praise to Allah (SWT) for keeping me well and having the opportunity of higher education.

I am thankful to my supervisor Prof. Steven G. Gilmour for his continuous support and pragmatic and relaxed supervision that has a great similarity with that caring of my parents who never forced me to do anything after my schooling year 4. I am grateful to Dr. Lawrence Pettit who agreed to became my first supervisor officially when Prof. Gilmour left for University of Southampton. I am thankful to Dr. Lawrence for reviewing my thesis and taking care of me in absence of Prof. Gilmour. Thanks to my PhD examiners for their recommendations to improve the quality of the thesis after the first submission.

I am grateful to my wife Lisa and lovely daughter Zahra for their long patience and support throughout my studies.

I love Queen Mary including its staff for their continuous cordial support and also for funding me through the DTA scholarship during this study. I am thankful to Dr. Peter Zemroch and Pauline Ziman for advising me during ImpactQM knowledge transfer project at Shell Global Solutions UK. I am grateful to Prof. Peter Goos, University of Antwerp, Belgium, for his collaboration with us and allowing us to use data from

Belgian automobile industries.

I would like to acknowledge University of Dhaka for the approval of my study leave for the whole period of PhD research.

I am thankful to my students who became colleagues later particularly Iftakhar and Zakir, friends Shihan and Li to have great gossips along with teas, coffees, biscuits, custard doughnuts and so on within the School of Mathematical Sciences building regularly for the last couple of years. These mind refreshing events would never be forgotten for the rest of my life.

Last but not the least, thanks to my great parents who are endless source of my inspiration for encouraging me to keep going my academic studies up to the PhD level.

Abstract

Industrial experimental design is an important area under design of experiments and factorial design hold a firm place in industrial experiments. The generalization of factorial designs results in split-plot type designs when complete randomization of runs is not possible. More specifically, hard-to-set factors lead naturally to split-plot type designs and mixed models. Mixed models are used to analyze multi-stratum designs as each stratum may have a random effect on the responses. The study of random effects in mixed models might be difficult using likelihood methods because of small number of groups or whole plots in multi-stratum and split-plot designs. Also, zero estimates of variance components could be due to estimating multiple variance components in a hierarchical model. Therefore, likelihood-based inference is often unreliable with the variance components being particularly difficult to estimate for small samples. A Bayesian method considering some non-informative or weakly informative priors for variance components could be a useful tool to solve the problem.

Fuel economy experiments, conducted by Shell Global Solutions UK, fall under small sample trap during variance components estimation. Using SAS procedure MIXED, experimenters estimated the variance components to be zero which were unrealistic. Also, the experimenters were unsure about the parameter estimates obtained by likelihood method from linear mixed models. Therefore, we looked for an alternative to compare and found the Bayesian platform to be appropriate. Bayesian methods assuming some non-informative and weakly informative priors enable us to compare the parameter estimates and the variance components. Profile likelihood and bootstrap based methods verified that Bayesian point and interval estimates are not absurd. Also, simulation studies have assessed the quality of likelihood and Bayesian estimates in this study.

A polypropylene experiment was conducted by four Belgian automobile industries to look for economical plasma treatments which lead to a good adhesion to various coatings. The effects of several additives were also studied in addition to the plasma treatments. The likelihood-based estimates were not reliable completely due to the existence of moderate number of whole-plots. Also, some of the variance components due to batch were zero for some coatings. Assuming non-informative priors for fixed effects and some weakly informative priors for variance components we have obtained more sensible estimates of variance components which were inestimable or poorly estimated by the likelihood-based method using SAS procedure GLIMMIX. In this study, the Bayesian methods appeared to give comparable results with classical methods.

One response variable in the polypropylene experiment was categorical which was converted to binary to see the effects of additives on the outcome of interest. Unfortunately for binary responses we failed to obtain estimates of the logistic parameters for some of the coatings as the system did not converge. One of the reasons for this was due to having the separation problem in the data. When one or more explanatory variables completely separate the responses, the problem is known as separation. This problem causes the non-existence of likelihood estimates of logistic regression parameters.

We have done some novel methodological works on the separation issue to minimize the problem in the light of optimal design techniques. Though the information based D-optimality criterion is widely used in practice, it fails to handle the separation problem appropriately. We have proposed new probability-based optimality criteria to handle the separation problem at the design stage of a study. Our proposed criteria P_s- and DP_s- might be worthwhile to take into account reduction of the separation problem. However, P_s-criterion alone is not suitable to deal with separation problem as it produces worse designs in terms of precision of the parameter estimates, i.e. with respect to D-optimality. On

the other hand the compound DP_s-criterion makes a balance between D- and P_s-optimality and produces better designs. To make designs less sensitive to parameter misspecification, pseudo-Bayesian design criterion DP_{SB}- has been proposed. Simulation studies have verified that Bayesian designs perform better than non-Bayesian designs by providing less bias, less median squared errors and above all less probability of separation. Thus, newly devised Bayesian and non-Bayesian design criteria could be useful in practice to control separation problem at the design stage of a study.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Preface

Undertaking experiments is a natural way to realize the best way to do things. It is a common phenomenon to do experiments in any scientific discipline for striving towards perfection. Statistical experimental designs attempt to answer how, with a minimum of effort, one can discover which factors do what to which responses. In other words, experimental design or design of experiments is a structured and organized way to conduct and analyze controlled tests for evaluating the factors that affect response variables.

Design of experiments was pioneered by Ronald A. Fisher in the 1920s and early 1930s at Rothamsted Experimental Station, an agricultural research station 25 miles north of London. Fisher recognized that flaws in the way in which the experiment that generated the data had been performed often hampered the analysis of data from agricultural systems. With the interactions of scientists and researchers in various fields, he developed the insights that led to the three basic principles of experimental design: randomization, replication, and blocking. Randomization means random assignment of experimental units to the levels of a treatment. It helps distributing the unusual characteristics of experimental units over the treatment levels so that they do not selectively bias the outcome of the experiment. Random assignment also allows the computation of an unbiased estimate of error effects which are not attributable to the manipulation of the independent variable and it makes sure that the errors are statistically independent. Replication is the

observation of two or more experimental units under the same conditions. Replication is needed to estimate the variance of errors and obtain a more precise estimate of treatment effects. Blocking is an experimental procedure for isolating variation attributable to an extraneous factor. Blocking helps to remove the influence of extraneous factors as extraneous factors are undesired sources of variation that can affect the response variable [Montgomery, 2008].

Although the experimental design technique was first used in an agricultural context, the technique has been extended successfully in industry since the 1940s. The industrial experimental era was catalyzed by the introduction of response surface methodology (RSM) by Box and Wilson in 1951. They identified the fundamental differences between agricultural and industrial experiments. In industrial experiments, the response variable can be observed shortly after the experiment and experimenter can learn quickly important information from a small group of runs that might be utilized to plan the next experiment. The RSM technique was widely used in the chemical and process industries, particularly in research and development works. Taguchi et al. [1987] advocated for robust parameter designs specifically in making processes insensitive to environmental or other factors, obtaining products insensitive to variation transmitted by factors, and finding levels of the process variables that force the mean to a desired value while simultaneously reducing variability around this value [Montgomery, 1999]. However, Taguchi's methods were criticized widely as his methods were advocated primarily by entrepreneurs in the West and as the underlying statistical science had not been adequately peer reviewed [Montgomery, 2008]. Though Taguchi's methods were criticized, his efforts appeared to have positive impacts by instigating designed experiments in the discrete part industries including automotive and aerospace manufacturing, electronics and semiconductors, and many other industries that had little use of experimental design techniques.

Applications of designed experiments have grown substantially in industries. More or less industrial experimental design techniques have some common approaches to follow. The following steps are useful while one may be performing an industrial experiment.

1. Identification and statement of the problem.
2. Selection of the response variable.

3. Selection of the process variables, noise factors and the interactions among the process variables of interest.
4. Determination of factor levels and range of factor settings.
5. Choice of appropriate experimental design.
6. Experimental planning and performing the experiment.
7. Statistical data analysis, interpretation and recommendations.

The above points are self explanatory and yet further details on these are also available in Antony and Capon [1998] and Montgomery [2008].

Though experimental design techniques are very powerful but there are some problems associated to implementation of the techniques to industries. In many situations, there is lack of communication between the industrial and the academic worlds, therefore it limits the use of experimental design in many manufacturing industries. There is also lack of adequate skills and expertise required by engineers in manufacturing to define and formulate problems. Thus many engineers face difficulties in analysing a particular process quality problem and then converting the engineering problem into statistical terms from which appropriate solutions can be chosen. Further, even after accomplishing experiments, problems may arise from a statistical point of view, e.g. parameter estimates may not exist due to censoring in lifetesting experiments particularly connected to small experiments [Hamada and Tse, 1992]. Due to small number of units often estimates of some parameters, e.g. variance components in mixed models, might not be estimable reliably. Further, due to the nature of the data, computational obstacles may arise, for instance, maximum likelihood estimates (MLEs) may not exist for particular types of models under designed experiments. However, despite these deficiencies, experimental designs are increasingly advancing in industrial experiments as they catalyze scientific methods and greatly increase efficiency in industrial production in terms of less cost, less time and better quality.

In the field of industrial experimental designs factorial designs hold a firm place. A researcher selects a fixed number of levels of each of the factors in a factorial design and then runs the experiment with all possible combinations. If the researcher is unable to randomize completely the order of runs that results in generalizations

of the factorial design called split-plot designs, strip-split plot designs etc. Split-plot designs are widely used in industrial experiments. Complete randomization is not possible often in industrial experiments because some factors may have levels that are difficult to set. If hard-to-set factors are considered at the design stage, they lead naturally to multi-stratum structures, with different factors applied in different strata through restricted randomization, as in split-plot designs [Goos and Gilmour, 2012]. Further, many split-plot designs yield categorical response data. The combination of categorical data and restricted randomization necessitates the use of generalized linear mixed models as each stratum under multi-stratum designs may have random effects on the responses.

The study of random effects in mixed models is often difficult because of small number of groups (number of strata or whole plots in multi-stratum and split-plot designs), particularly difficulties with the estimation of variance components and consequently with the statistical inference. To avoid these difficulties of estimation, Bayesian analysis is suggested by earlier researchers for normal responses assuming some non-informative or weakly informative priors for the relevant parameters. This is extended to discrete responses in this Thesis.

During logistic regression analysis of binary or categorical response data in experiments, researchers often face convergence difficulties due to a problem known as 'separation'. It is an undesirable problem in models for dichotomous dependent variables. It occurs when one or more model covariates perfectly predicts some binary outcome. Current literature suggests that researchers need to compromise with separation either by undertaking *post hoc* data adjustment or by estimation corrections. However, apart from these solutions, the separation issue could be handled by using newly developed optimality criteria at the design stage of an experiment.

1.2 Problems Addressed in the Thesis

There are two main issues in this Thesis. Firstly, estimation of variance components that were poorly estimated or estimated as zero in multi-stratum industrial experiments. Secondly, handling separation problem that causes infinite estimates

or non-existence of maximum likelihood estimates (MLEs) of parameters in logistic regression model. For estimating the variance components we have implemented Bayesian techniques as an alternative to likelihood based methods and for separation problem we have used an optimal design technique to avoid separation at the design stage of a study. However, the two things - non-existence of MLE (or infinite estimates) of parameters and zero estimates of variance components should not be confused and should be treated as two separate issues- former is about random effects and latter is about fixed effects. Bayesian methods implemented here are not to show outright domination over frequentist methods, rather we consider as complementary to each other and where frequentist (likelihood) methods fail then Bayesian methods appeared to assist or vice versa.

Increasingly analysis becomes cumbersome for complex models. Often researchers want to compare results by applying several methods. The likelihood estimates of variance components were zero in fuel economy and polypropylene industrial experiments. The main reason of having zero estimates is small to moderate number of experimental units. Also Bayarri and Berger [2004] have clarified that maximum likelihood estimates of variance components in hierarchical models (or variance components models) can easily be zero, especially when there are several variance components in the model that are being estimated. It is also quite common in problems with numerous variance components to have at least one MLE variance estimate equal to zero. Therefore, as the likelihood (frequentist) method fails we consider Bayesian method as an alternative to obtain realistic estimates of variance components. But appropriate choice of priors is crucial in Bayesian analysis. Bayarri and Berger [2004] noted that frequentists are usually not interested in subjective, informative priors, and on the other hand Bayesians are less likely to be interested in frequentist evaluations when using subjective, highly informative priors. Again the most common scenarios of useful connections between frequentists and Bayesians are when no external information (other than the data and the model itself) are introduced into the analysis - in the Bayesian context, when "objective prior distributions" are used. As we do not have firm basis to consider any informative priors, we have implemented non-informative or weakly informative priors throughout our analysis of fuel economy and polypropylene experiments to keep things close to frequentist ideas. However, performance of the Bayesian and the likelihood based estimates has been assessed through simulation studies.

We have conducted a Bayesian analysis of data from a multi-stratum design with discrete responses in polypropylene experiment. The originality of the work is to adapt the methodology to a new application area. On the basis of this experience from the Bayesian study of polypropylene experiment, we had an opportunity to work for Shell Global Solutions UK, an energy consultant and technology innovator, through a knowledge transfer project called 'ImpactQM Shell Transfer Project' that produces the second chapter in the Thesis. This is a pioneering research where we have applied Bayesian methods in the fuel efficiency field, which has also a particular feature of using very small experiments. This method is now recommended in the Shell industrial guidelines. Finally, a statistical computational problem known as 'separation' arises during binary response analysis under logistic regression that leads us do some novel methodological works on optimal design of experiments considering separation issues. The rest of this section elaborates slightly on the topics that we studied.

In a fuel economy experiment, researchers wanted to analyze data in different ways as they were unsure which methods to adopt in practice. This experiment was run as a multi-stratum design as there were several strata (sessions-morning vs afternoon, days nested under weeks) in the experiment. Initially they applied likelihood-based methods, particularly linear mixed models to model the continuous responses. The purpose was to compare performances of car fuels. However, variance components due to random effects were estimated as 0 which was unrealistic. In their further investigation, they found the Bayesian method as a possible alternative described in Gilmour and Goos [2009] to estimate the variance components. Assuming some weakly informative priors on the variance components, the problem can be resolved, thus estimating the parameters and their standard errors more precisely. Simulation studies were accomplished to determine which method likelihood or Bayesian produces better estimates and further which set of priors produces the best estimates. The quality of point estimates was assessed by bias, relative bias, mean and median square errors and quality of interval estimates was measured by the coverage probabilities, average and median widths of the confidence/credible intervals.

Generalized linear models (GLMs) are widely used to analyze categorical response data, including in factorial designs. The combination of categorical data and re-

stricted randomization in multi-stratum designs necessitates the use of generalized linear mixed models. The successful estimation of parameters and variance components of a mixed model depends on the number of whole plots in a split-plot design. An inadequate number of whole plots is a hindrance in the proper estimation of variance components. In the polypropylene industrial experiment, the number of whole plots (batches) was moderate, therefore, the experimenters could not estimate a variance component accurately [Goos and Gilmour, 2012]. However, a Bayesian method assuming some weakly informative priors for variance components was suggested to overcome the hurdle. In this study, we have implemented the Bayesian techniques to estimate the variance components that were inestimable or poorly estimated in likelihood-based methods.

When responses are binary, then often it might be impossible to obtain maximum likelihood estimates (MLEs) of the logistic model parameters due to convergence difficulties. This happens due to a problem known as 'separation' that occurs due to the perfect prediction of outcome of interest by one or more covariates. It is a phenomenon associated with models including logistic and probit models for binary or categorical outcomes. We have carried out a study on the separation issue in the context of optimal design theory.

1.3 Literature Review

In this section, some literature have been reviewed that will help understanding the background of this study.

1.3.1 Bayesian Analysis of Data from Multi-Stratum and Split-plot Designs

In some multi-factor experiments complete randomization is not feasible. This often results in a generalization of the factorial design called the split-plot designs. A split-plot design is a blocked experiment, where blocks themselves serve as experimental units for a subset of the factors. The blocks are referred to as whole plots, while the experimental units within blocks are called split plots, split units or subplots [Jones and Nachtsheim, 2009]. Hard-to-set factors in split-plot designs lead

naturally to multi-stratum structures if they are taken into account at the design stage. Mixed models are used to analyze multi-stratum designs as each stratum may have random effects on the responses.

Letsinger et al. [1996] pointed out that the results from data analysis of split-plot response surface designs could be misleading if experimenters ignore the multi-stratum structure. For normal responses, they recommended the use of a linear mixed model with variance component parameters estimated by residual maximum likelihood (REML) and fixed effect parameters by empirical generalized least squares (GLS). This has become the standard way of analysing data from industrial multi-stratum experiments and is usually successful if there are ample residual degrees of freedom in each stratum.

A disadvantage of REML-GLS estimation is that it can give highly undesirable and misleading conclusions in non-orthogonal split-plot designs with few main plots. This proved to be true in a freeze-drying coffee experiment reported by Gilmour et al. [2000] where the main plot variance component was estimated to be 0 using REML-GLS methods. Experience suggested that this was implausible, but inferences and estimated standard errors for fixed effects use this estimate. Gilmour and Goos [2009] implemented a Bayesian method using informative priors for the main plot variance components in linear mixed models (LMMs) for the freeze-drying coffee experiments where the responses were normal.

Industrial experiments frequently involve non-normal response data as in other areas of application. Examples of non-normal response could be a binary response e.g. defective or non-defective, success or failure and so on, the response of interest can be a count, e.g. the number of faults in an item or number of success in an experiment. Goos and Gilmour [2012] analyzed binary and categorical data in a multi-stratum polypropylene experiment using generalized linear mixed models and a likelihood-based estimation and inference approach implemented in the SAS procedure GLIMMIX. Some of the variance components were estimated to be 0, perhaps due to the small number of plots in the higher strata or due to estimating several variance components simultaneously as clarified by Bayarri and Berger [2004]. Therefore, they suggested the possibility of performing a Bayesian analysis assuming weakly informative priors for the variance components corresponding

to the higher strata. Fong et al. [2010] also mention that likelihood-based inferences can be unreliable, with variance components being particularly difficult to estimate for small samples. In this study, we have implemented the Bayesian methods assuming some informative and weakly informative priors for the variance components and obtained reasonable estimates of variance components that were inestimable or poorly estimated by likelihood-based methods.

1.3.2 Non-existence of Maximum Likelihood Estimates and Separation Problem in Logistic Regression

During analysis of categorical data under various models, for instance, Poisson, binomial, multinomial, log-linear models including logit models, often maximum likelihood estimates (MLEs) do not exist. Fienberg and Rinaldo [2007] used the wording "non-existence of the MLE" to signify lack of solutions for the maximum likelihood optimization problem. The reason behind this is the existence of the sampling zeros in the contingency tables. The non-existence of MLE is not only dependent on zeros in contingency table but also the position of zeros in the table. Being caused by the presence of sampling zeros, non-existence of the MLE is more likely to occur in sparse tables with small counts, a setting in which the traditional χ^2- asymptotic approximation to various measures of goodness of fit is known to be unreliable. Specific examples how sampling zeros, where separation is a special case, causes non-existence of MLEs have been given in Chapter 4. Also further details on the non-existence of MLE and the position of sampling zeros in the contingency table are given in Fienberg and Rinaldo [2007]. Haberman [1974] discusses MLE non-existence for log-linear models where logistic model is a special case. In his terminology, MLE existence means finiteness of the solution. He proves a very general theorem on necessary and sufficient conditions for the maximum likelihood estimate to exist and also he demonstrates that for most models, if the maximum likelihood solution exists, it is unique, as a result of the concavity of the likelihood function. Necessary and sufficient conditions for existence constitute a linear programme, which is typically hard to solve in practice. Silvapulle and Burridge [1986] and Hamada and Tse [1988] showed that for popular reliability models, such as log-normal, Weibull, and exponential regression models, the question of the MLEs existence reduces to solving a LP problem. For simple linear regression, Hamada and Tse [1988] described how the LP problem can be reduced to checking a few data configurations. However, for small industrial factorial ex-

periments Hamada and Tse [1992] shows that a simple linear programming (LP) problem can be solved by a standard LP algorithm.

All problems of non-existence of the MLE depend on the positions of the sampling zeros in the contingency table. It is observed that not all possible sampling zeros are causing non-existence problem. Even with the existence of positive margins there can occur MLE non-existence problem due to having some zeros inside contingency table. Although non-existence of the MLE arises most frequently in sparse tables, it can very well occur in tables with large counts and very few zero cells [Fienberg and Rinaldo, 2007].

Fienberg and Rinaldo [2007] describes the implications of sampling zeros for the existence of maximum likelihood estimates for log-linear models. Understanding the problem of non-existence is crucial to the analysis of large sparse contingency tables. Gloneck et al. [1988] proved that positivity of the margins is a necessary conditions for existence of the MLE if and only if the model is decomposable. Specific examples on how sampling zeros in the contingency table can cause MLE non-existence are have been given in Section 4.2 of Chapter 4. One special case of sampling zeros in the contingency table is separation problem where response is completely separated into two parts. Separation causes numerical problem of non-existence of MLE for logistic model.

Haberman [1977] discusses likelihood equations and necessary and sufficient conditions in exponential models, where exponential response models are the generalization of logit models for quantal responses. He also explores the asymptotic properties of MLEs.

As mentioned before, although Haberman [1974] gave necessary and sufficient conditions for the existence of the MLE in log-linear models including logistic model, his characterization is non-constructive in the sense that it does not directly lead to implementable numerical procedures and also fail to suggest alternative methods of inference in case of an undefined MLE. Despite these deficiencies, Haberman's (1974) results have remained all that exist in the published statistical literature [Fienberg and Rinaldo, 2007]. Perhaps this is the main reason that surprisingly the authors, for instance, Heinze and Schemper [2002], Bryson and Johnson

[1981], Santner and Duffy [1986], Silvapulle [1981], Zorn [2005], Clarkson and Jennrich [1991], Jacobsen [1989], Altman et al. [2004], Allison [2008], Agresti and Yang [1987], Lamotte [2005] who deal with separation issue, did not mention Haberman [1974]'s work in their literature. Again, Albert and Anderson [1984] terms that Haberman [1974]'s works rather theoretical and has not been useful in solving real life problems related to non-existence of MLE. Therefore, it is not astonishing that virtually all implemented computational algorithms for fitting log-linear models are incapable of handling these cases stated by Haberman [1974]. For example, in SAS, the presence of sampling zero is dealt with by adding small positive quantities to the zero cells to facilitate the convergence of numerical procedures. However, this common practice can be misleading as clarified by Fienberg and Rinaldo [2007].

Furthermore, no one has presented yet a numerical procedure specially designed to check the existence of the MLE and the only indication of non-existence is lack of convergence of whether algorithm is used to compute the MLE. As a result, the possibility of non-existence of the MLE, even though well known, is rarely concern for the practitioners. Moreover, even for those cases in which the non-existence is easily detectable e.g. when the observed table exhibits zero margins, there do not exist appropriate inferential procedures for dealing with such situations [Fienberg and Rinaldo, 2007].

The event of separation is observed during the fitting process of logistic models where at least one parameter estimate diverges to $\pm\infty$. Separation primarily occurs in small samples with several unbalanced and highly predictive covariates. The name 'separation' is given by Albert and Anderson [1984] as responses and non-responses are perfectly separated by a single factor or by a non-trivial linear combination of factors [Heinze and Schemper, 2002]. The phenomenon of separation is also known as 'monotone likelihood' [Bryson and Johnson, 1981]. The problem of separation is by no means negligible and may occur even if the underlying model parameters are low in absolute value. Substantively, separation often forces researchers to make difficult, consequential, and largely arbitrary choices about data, measurement, and model specification [Zorn, 2005].

Logistic regression data sets were classified by Albert and Anderson [1984] into three mutually exclusive and exhaustive categories namely complete separation,

quasi-complete separation, and overlap. To check whether the data is of type complete, quasi-complete, or overlapped Santner and Duffy [1986] developed a linear programme. They verified that the maximum likelihood estimates exists only for the overlapped data. Clarkson and Jennrich [1991] also developed similar algorithms to compute extended maximum likelihood estimates when one or more parameter estimates are infinite at the supremum of the likelihood.

Often in medical and other research, the outcome is binary and parameter estimates of logistic regression are not available (see examples in Heinze and Schemper [2002] and Silvapulle [1981]). In general, one does not assume infinite parameter values in underlying populations. The problem of separation is rather one of non-existence of the maximum likelihood estimate under special conditions in a sample [Jacobsen, 1989]. An infinite estimate can also be regarded as extremely inaccurate, the inaccuracy resulting in Wald confidence intervals of infinite width [Lesaffre and Marx, 1993].

Altman et al. [2004] and Allison [2008] explained how and why numerical algorithms for maximum likelihood estimation of the logistic regression model sometimes fail to converge due to separation. Heinze and Schemper [2002] have shown how the probability of separation depends on sample size, on the number of dichotomous covariates, the magnitude of the odds ratios associated with them and on the degree of balance in their distribution.

The solutions of separations are addressed in many ways in the literature. The naive method is the deletion of the variable(s) causing separation. Omission of the problem variable(s) is strongly discouraged as no information remains available on it though it might be an important factor. Deleting variables with strong effects will certainly obscure the effects of those varibles and is likely to bias the coefficients for the other variables in the model Allison [2008].

Ad hoc adjustment (data manipulation) prior to a standard analysis may produce finite estimates. However, simple adjustment of cell frequencies can have undesirable properties [Agresti and Yang, 1987]. Researchers often choose a different model instead of the logistic regression model to fit the available data. Changing to a different model might help to avoid the problem, but models whose parameters

have different interpretations that are not relevant to the logistic model may be less appealing.

Hirji et al. [1989] demonstrated that the use of exact logistic regression allows replacement of unsuitable maximum likelihood estimates by a median unbiased estimate. However, this method may be unsuitable with the existence of a continuous covariate or multiple dichotomous covariates [Heinze and Schemper, 2002].

Firth [1993] developed penalized maximum likelihood estimation (PMLE) by a modification of the score function of logistic regression to reduce the bias of MLEs. However, Wald tests based on the standard errors for variables causing separation can be highly inaccurate similar to other conventional ML methods.

Lamotte [2005] bypasses the issue of separation by finding the supremum of likelihood function of the response variable and thereby computed exact conditional p-values based on the likelihood ratio statistic in logistic regression. However, there is no indication how to find intermediate probabilities of outcome variables given the covariates by his method.

In all of these studies discussed above, the emphasis is mainly on distinguishing whether MLEs exist for a single outcome or, when they do not, finding reasonable substitutes, or adjusting other things to bypass separation issues. None of these studies addresses the problem in the light of optimal design of experiments. The problem can be controlled at the design stage of an experiment to reduce the risk of facing the problem of separation at the estimation stage. We propose new probability-based optimality criteria that will be minimized to generate optimal design values that will reduce the probability of separation during analysis of experiments and thus enable estimation of MLEs appropriately. Neither forcefully omitting of any covarite or *post hoc* adjustment will be required when applying the method of probability-based optimal designs.

1.4 Structure of Thesis

The rest of thesis is organized as follows. In Chapter 2, the Bayesian analyses of fuel economy experiments are described under multi-stratum designs. The Bayesian analyses of data from multi-stratum and split-plot designs with discrete responses are discussed in Chapter 3. Some novel methodological works have been accomplished concerning optimal design of experiments with separation under logistic regression in Chapter 4. Finally, some discussions and conclusions have been drawn in Chapter 5. A sample of computer codes written in WinBUGS and the R statistical programming language is found in an Appendix.

Chapter 2

Analysis of Fuel Economy Experiments Using Bayesian Methods

2.1 Introduction

In the context of transport, fuel economy refers to the fuel efficiency relationship between the distance travelled by an automobile and the amount of fuel consumed. Fuel economy is expressed in miles per gallon (mpg) or kilometres per litre (km/L). Fuel efficiency is dependent on many aspects of a vehicle, for instance, engine parameters, aerodynamic drag, weight, rolling resistance. To improve economic usage of fuel, many types of experiment are being done in automobile laboratories.

Shell Global Solutions UK, a leading-edge energy consultant and technology innovator, has conducted many experiments to distinguish performances of different fuels. The data from one of the fuel economy experiments were analyzed by Shell using classical methods. As the experiment was expensive, the experimenters were interested to analyze the data using some other statistical methods for comparison purposes. However, more specifically, this work was motivated by the estimation problems of variance components in fuel economy and round robin experiments. The classical methods estimated variance components to be zero which was not realistic. A Bayesian approach may resolve the problem associated with the zero estimates of variance components by introducing a certain amount of prior in-

formation on the parameters. Therefore, a Bayesian method was implemented to compare with the outputs obtained from the likelihood method as well as to overcome the difficulties related to the variance component estimation. Further, simulation studies were carried out to assess, and to compare the quality of point and interval estimates obtained from likelihood and Bayesian methods.

This study was a part of a knowledge transfer project known as 'ImpactQM-Shell Transfer Project' where I have implemented Bayesian methods under joint supervision of my PhD adviser and experts based at Shell Technology Centre, Thornton, Chester, UK.

2.2 Bayesian Models

A statistical model describes the relationship between variables in the form of mathematical equations. Let us, for example, consider a simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{2.1}$$

where $Y_i \sim N(\mu_i, \sigma^2)$, with $\mu_i = \beta_0 + \beta_1 X_i$ and $\tau = 1/\sigma^2$ (a precision parameter), β_1 is the rate of change in $E(Y)$ due to change in X and ϵ is an error term. In classical method the parameters (i.e. β_0 and β_1) are assumed to be fixed. In Bayesian analysis β_0 and β_1 also follow some distributions. This means that some prior information is available on β_0 and β_1. The prior distributions of β_0 and β_1 are subjective. In Bayesian models we change the form of normal distributions from $N(\mu, \sigma^2)$ to $N(\mu, \tau)$, where $\tau = 1/\sigma^2$. Therefore, from now onwards, we will use the precision parameter (τ) instead of variance parameter (σ^2) in the Bayesian models. The estimates obtained combining prior belief and available data are called posterior estimates.

Bayesian inference determines the posterior distribution of the parameter using

Bayes' theorem. A Bayesian version of model (2.1), for example, can be written as

$$\begin{aligned} Y_i &\sim \mathrm{N}(\mu_i, \tau) && (2.2)\\ \beta_0 &\sim \mathrm{N}(0, 0.0001) && (2.3)\\ \beta_1 &\sim \mathrm{N}(0, 0.0001) && (2.4)\\ \tau &\sim \mathrm{Gamma}(0.0001, 0.0001)\,. && (2.5) \end{aligned}$$

It is found in classical methods that the estimates of β_0 and β_1 follow normal distributions for large samples. However, the estimates of β_0 and β_1 follow normal distributions even with small samples as long as the normality assumption holds for the residuals (ϵ) [Searle, 2012]. $\beta_0 \sim \mathrm{N}(0, 0.0001)$ means that β_0 follows a normal distribution with mean 0 and variance 10000. Often gamma priors are assumed for precision parameters. $\tau \sim \mathrm{Gamma}(0.0001, 0.0001)$ means that τ follows a gamma distribution with mean 1 and variance 10000. In equations (2.3), (2.4) and (2.5) we have used noninformative priors for β_0 , β_1 and τ. The rationale for using noninformative prior distributions is often said to be to let the data speak for themselves, so that inferences are unaffected by information external to the current data [Palta, 2003].

2.3 Bayesian Inference

Let D denote observed data, and θ denote model parameters. The joint distribution of D and θ is $P(D, \theta)$ which can be expressed as

$$P(D, \theta) = P(D|\theta)P(\theta), \tag{2.6}$$

where $P(D|\theta)$ is a likelihood and $P(\theta)$ is a prior.

Having observed D, Bayes theorem is used to determine the distribution of θ conditional on D

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}. \tag{2.7}$$

$P(\theta|D)$ is called the posterior distribution of θ, and is the object of all Bayesian

inference. $P(D)$ is called the **marginal distribution** of D. $\int P(D|\theta)P(\theta)d\theta$ is called the **normalizing constant**. The posterior expectation of θ is

$$E[\theta|D] = \frac{\int \theta P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}. \tag{2.8}$$

Until recently the integrations in (2.7) and (2.8) have been the source of most of the practical difficulties in Bayesian inference, especially in high dimensions. In most applications, analytic evaluation of the expectation (population mean) $E[\theta|D]$ is impossible. Alternative approaches of evaluations are numerical approximation, Laplace approximation, and Markov chain Monte Carlo (MCMC) integration. There is no doubt that the introduction of Markov chain Monte Carlo methods has revolutionized Bayesian statistics [Besag, 2001].

2.4 Markov Chain Monte Carlo (MCMC)

2.4.1 Why MCMC in Bayesian Methods?

The ability to integrate complex and high dimensional functions is extremely important in Bayesian statistics, whether it is for calculating the normalizing constant, the marginal distribution, or the expectation. Often, an explicit evaluation of the integrals, for instance the integrals defined in (2.7) and (2.8), is not possible for higher dimensions and, traditionally, we would be forced to use numerical integration or analytic approximation techniques [Brooks, 1998]. The Markov chain Monte Carlo (MCMC) method provides an alternative whereby we sample from the posterior directly, and obtain sample estimates of the quantities of interest, thereby performing the integration implicitly.

2.4.2 Three Related Terms

Markov chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain.

The **Metropolis-Hastings** algorithm is a MCMC technique for obtaining a sequence of random samples from probability distributions for which direct sampling

is difficult. This sequence can be used to approximate the distribution (i.e. to generate a histogram) or, to compute an integral (such as an expected value).

Gibbs Sampling is a special case of the **Metropolis-Hastings** algorithm which is usually faster and easier to use but is less generally applicable. This algorithm is used to generate a sequence of samples from a joint distribution of two or more random variables.

2.4.3 Gibbs Sampling

The technique of Gibbs sampling has been applied through a widely used Bayesian platform WinBUGS in all analyses of the current thesis. Gibbs sampling is one of the MCMC methods. The basic idea of MCMC methods is to simulate from the distribution we are interested in and then use the simulated sample to estimate parameters.

Suppose we have a problem with parameters, say β_0, β_1, and β_2. Gibbs sampling works by iteratively drawing samples from the full conditional distributions of unobserved nodes. The algorithm proceeds according to the following steps:

Step 1: Choose initial estimates $\beta_0^{(0)}$, $\beta_1^{(0)}$, and $\beta_2^{(0)}$. Let $j = 0$.

Step 2: Given current estimates $\beta_0^{(j)}$, $\beta_1^{(j)}$, and $\beta_2^{(j)}$ simulate new values
$\beta_0^{(j+1)}$ from $P\left(\beta_0|\beta_1^{(j)}, \beta_2^{(j)}, D\right)$
$\beta_1^{(j+1)}$ from $P\left(\beta_1|\beta_0^{(j+1)}, \beta_2^{(j)}, D\right)$
$\beta_2^{(j+1)}$ from $P\left(\beta_2|\beta_0^{(j+1)}, \beta_1^{(j+1)}, D\right)$

Step 3: Put $j = j + 1$ and return to Step 2.

Posterior Estimates

The posterior means of β_0, β_1, and β_2 can be estimated as follows

$$\hat{\beta}_0 \approx \frac{1}{N}\sum_{j=1}^{N}\beta_0^{(j)}, \quad \hat{\beta}_1 \approx \frac{1}{N}\sum_{j=1}^{N}\beta_1^{(j)}, \quad \hat{\beta}_2 \approx \frac{1}{N}\sum_{j=1}^{N}\beta_2^{(j)},$$

where N is the sample size, i.e. the number of MCMC samples used. It is essential to throw away some iterations at the beginning of an MCMC run as in the early time steps the probability distribution is not straight away like the target distribution. The influence of the arbitrary starting points is not desired. The early period that is excluded is known as the 'burn-in' period [Kruschke, 2011]. If the burn-in period has length M then estimates will be as follows

$$\hat{\beta}_0 \approx \frac{1}{N-M}\sum_{j=M+1}^{N}\beta_0^{(j)}, \quad \hat{\beta}_1 \approx \frac{1}{N-M}\sum_{j=M+1}^{N}\beta_1^{(j)},$$

$$\hat{\beta}_2 \approx \frac{1}{N-M}\sum_{j=M+1}^{N}\beta_2^{(j)}.$$

In the burn-in period the first M iterations are removed from the sample in order to avoid the influence of the initial values. If the generated sample is large enough, the effect of this period on the posterior estimates is minimal [Ntzoufras, 2009].

2.4.4 Software to Implement MCMC

Though MCMC methods are used widely in the Bayesian statistical community, interestingly few programmes are available for their implementation. This is partly because algorithms are generally fairly problem specific and there is no automatic mechanism for choosing the best implementation procedure for any particular problem [Brooks, 1998]. However, BUGS (Bayesian inference Using Gibbs Sampling) has appeared to solve some of these problems and is widely used by statistical practitioners. The Windows version of the BUGS programme is known as WinBUGS which is a freeware package (see http://www.mrc-bsu.cam.ac.uk/bugs/ and Lunn et al. [2000]). In addition to this, recently available other open source softwares are Bayesian Filtering Library (BFL), Just another Gibbs sampler (JAGS), LaplacesDemon, GNU MCSim, and Stan. However, we will use two freeware soft-

ware namely WinBUGS and R throughout our studies.

WinBUGS is a programming language based software that is used to generate a random sample from the posterior distribution of the parameters of a Bayesian model [Ntzoufras, 2009]. The user only needs to specify data, the structure of the model under consideration, and some initial values for the model parameters.

2.5 Case Studies

We have investigated two anonymous fuels - test (T) and base (B). The data was anonymised by Shell as they are commercially important and artificially manipulated keeping them as realistic as possible. We have compared the performances of T and B including several contrasts. Nested models were studied to see the effect of factors under nesting. Round robin programmes were implemented in order to understand and quantify variation in test methods.

2.5.1 Fuel Economy Experiments

Fuels B (Base) and T (Test) are tested in order to assess which gives the better fuel economy in a vehicle. The experiment was run on three separate cars for two weeks each of which had three days. There were two sessions-morning and afternoon in each day. However, we will present analysis for one car only in this chapter. The response variable, measured on a continuous scale, was distance crossed by a car per gallon of fuel burned.

Underlying Design

Out of three cars, we consider the experiment with one car to explain the underlying design. Both the test and base fuels were tested in the car for two weeks each of which has three days (see Table 2.1). Each of the cars had two test sessions-morning and afternoon in a day. Once B or T is treated in a car, it is difficult to remove it from the fuel tank. The car had to undergo three back-to-back tests (BBB or TTT) in a session. In the morning (AM) of day 1 under week 1, there were three back-to-back tests on the base fuel (BBB) and similarly there were five

Table 2.1: Underlying design in the fuel economy experiment

Week	Day	Session	Tests
1	1	AM	BBB
		PM	BBB
	Overnight stand	-	-
	2	AM	BBB
		PM	BBB
	Overnight stand	-	-
	3	AM	BBB
		PM	BBB
INTERVAL OF 4 - 5 DAYS		FULL CLEAR OF TREATMENTS	
2	1	AM	BBB
		PM	BBB
	Overnight stand		
	2	AM	BBB
		PM	TTT
	Overnight stand	-	-
	3	AM	TTT
		PM	TTT

more series of back-to-back tests on fuel B in the first week. At the end of first week, there was an interval of 4-5 days before starting second round of the tests on the same car. In the meantime there was cleaning of treatments in the vehicle to remove previous fuel effects (if there were any) in the experiment. In the second week, the last three series of tests were on fuel T.

In the design B-T means that change of treatment from base fuel to test fuel and B-B or T-T means dummy change of treatment, implying actually no change of treatment. Some vehicles have a control week followed by a test week and others have a test week followed by a control week. We assume that the three back-to-

back tests in each half-day session have been averaged. We thus have two results per car per day, one from the morning and one from the afternoon. Variance components are considered for a treatment change, day to day variation or week to week variation subject to full clear out of treatment and break of 4-5 days.

2.5.1.1 Example Data Set

The fuel economy raw data set is presented in Table 2.2. The experiment was conducted in two weeks shown in the first column. Days were numbered as 1, 2, 3 for the first week and 4, 5, 6 for the second week. There were two sessions- morning and afternoon containing three trials each (e.g. BBB or TTT). The response variable Y on a continuous scale represents miles crossed by a vehicle given a unit gallon of fuel.

For simplicity, we assume that the three back-to-back tests in each half-day session have been averaged. We thus have two results per car per day, one from the morning and one from the afternoon. In Table 2.2 we treat half-day averages from all six days as single data point and calculate pooled estimates of between and within day variation. After manipulation, Table 2.2 has been summarized in Table 2.3 to study different contrasts and nested models.

The objective is to estimate the mean for each fuel and the difference in means and to find the median, 25th percentile, and 75th percentile of the posterior distributions of base fuel, test fuel, fuel difference, and day to day variance component.

2.5.1.2 Contrast: T-B

We extract data of week 2 from Table 2.3 to prepare Table 2.4 and on days 1-3 we compare fuels B and T combining between day and within day information.

Model

The mixed model (2.9) has been considered to analyze contrast T-B.

$$Y_{jkm} = \alpha + \beta_j + \delta_k + \epsilon_{jkm} \tag{2.9}$$

where $Y_{jkm} \sim N(\mu_{jk}, \tau)$ is the response corresponding to the m-th test ($m = 1$ or

Table 2.2: Data before averaging over back-to-back tests

Week	Day	Session	Treatment	Y
1	1	am	B	31.90993
1	1	am	B	31.61670
1	1	am	B	32.07328
1	1	pm	B	32.38294
1	1	pm	B	32.35951
1	1	pm	B	32.51994
1	2	am	B	31.92975
1	2	am	B	32.32851
1	2	am	B	31.67399
1	2	pm	B	31.79294
1	2	pm	B	31.49287
1	2	pm	B	31.44593
1	3	am	B	31.76795
1	3	am	B	31.36462
1	3	am	B	31.82962
1	3	pm	B	31.87879
1	3	pm	B	32.04046
1	3	pm	B	31.89035
2	4	am	B	32.11118
2	4	am	B	32.41172
2	4	am	B	32.43854
2	4	pm	B	32.08281
2	4	pm	B	32.60450
2	4	pm	B	32.17017
2	5	am	B	32.08908
2	5	am	B	32.15086
2	5	am	B	31.69741
2	5	pm	T	33.87101
2	5	pm	T	33.24747
2	5	pm	T	33.71225
2	6	am	T	33.08393
2	6	am	T	33.62343
2	6	am	T	33.31008
2	6	pm	T	33.86173
2	6	pm	T	33.89393
2	6	pm	T	33.14460

Table 2.3: Data averaged over back-to-back repeats

Week	Day	Session	Treatment	Y
1	1	am	B	31.86663
1	1	pm	B	32.42080
1	2	am	B	31.97741
1	2	pm	B	31.57725
1	3	am	B	31.65406
1	3	pm	B	31.93653
2	1	am	B	32.32048
2	1	pm	B	32.28583
2	2	am	B	31.97912
2	2	pm	T	33.61024
2	3	am	T	33.33915
2	3	pm	T	33.63342

2) on day k ($k = 1, 2, 3$) corresponding to the test fuel j ($j = 1$ or 2) with mean $\mu_{jk} = E(Y_{jkm}|\delta_k) = \alpha + \beta_j + \delta_k$ and precision $\tau = 1/\sigma^2$, β_j is the fixed effect due to the jth fuel with the constraint $\beta_1 = 0$, δ_k is the day-to-day error term (i.e. random effects due to kth day) which follows a normal distribution with mean zero and variance σ_b^2, and ϵ_{jkm} is the within-day error term with mean zero and variance σ^2. We can define within-day correlation by $\rho = \frac{\sigma_b^2}{\sigma_b^2+\sigma^2}$ or equivalently $\rho = \frac{\tau}{\eta+\tau}$ which can be also be expressed as $\eta = \frac{(1-\rho)\times\tau}{\rho}$, where $\eta = 1/\sigma_b^2$.

Table 2.4: Data to test contrast T-B

Week	Day	Session	Treatment	Y
2	1	am	B	32.32048
2	1	pm	B	32.28583
2	2	am	B	31.97912
2	2	pm	T	33.61024
2	3	am	T	33.33915
2	3	pm	T	33.63342

We perform Bayesian analysis for the contrast T-B assuming the following priors and using WinBUGS 1.4 (see the Appendix for WinBUGS codes related to this analysis).

Priors and Results

We assume priors corresponding to model (2.9) as follows

$$\alpha \sim N(38, 0.1), \quad \beta_j \sim N(0, 0.001)$$

$$\rho \sim \text{beta}(1, 1), \quad \log(\sigma) \sim U(-20, 20)$$

The prior for α was centered at 38 to incorporate the notion of mean fuel effect which might hover around 38 as believed by the Shell experimenters. Therefore, we assume a weakly informative prior for α by taking $\alpha \sim N(38, 0.1)$ and implemented at the beginning. However, Congdon [2007] suggested that, in the absence of prior information about the direction or magnitude of covariate effects, flat priors may be used by taking univariate normal distributions with mean zero and large variance. The effect of using normal priors with means 0 and large variances is that parameter estimates are smoothed towards zero as large variances are used [Galindo-Garre et al., 2004]. Therefore, we tried a non-informative prior for β_j by setting $\beta_2 \sim \text{N}(0, 0.001)$ i.e. β_2 follows a normal distribution with mean 0 and little precision 0.001 or large variance 1000. Also, we assumed a non-informative prior for σ by assuming $\log(\sigma) \sim U(-20, 20)$. The prior for the intra-class correlation ρ is non-informative which is used to compute the precision of the day-to-day error term. The results are presented in Table 2.5 assuming the above set of priors. However, a completely non-informative prior for α, for instance $\alpha \sim N(0, 0.001)$, and weakly informative prior $\alpha \sim N(38, 0.1)$ provide similar results to those presented in Table 2.5 as the concept of non-informative is originated from the assumption of very little precision.

There were some autocorrelation effects in the results before thinning (where thinning refers to removal of some values from the chain). When data were thinned by 15 (i.e. instead of using every step in the chain, we only used every 15th step), then the autocorrelation disappeared (see also the discussion in Section 2.6). Posterior means were calculated on the basis of the sample with size 10000. First, 1000 samples were ignored to remove initial fluctuations of the chains. The effect of the base fuel (B) is 32.17 mile/gallon and of the test fuel (T) is 33.59. The difference of effects (β_2) between test and base fuel is 1.421 with 95% Bayesian credible interval (0.952 , 1.877). In the table P(2.5) and P(97.5) denote the 2.5th and 97.5th percentiles of posterior estimates respectively. The variance between days (σ_b^2) is

Table 2.5: Results for contrast T-B

Effect	Mean	SD	MC Error	P(2.5)	Median	P(97.5)
B (α)	32.170	0.205	0.002	31.780	32.170	32.580
Fuel Diff(β_2)	1.421	0.240	0.002	0.952	1.420	1.877
T	33.590	0.209	0.002	33.190	33.580	34.040
μ_{11}	32.250	0.139	0.002	31.980	32.250	32.530
μ_{21}	32.250	0.139	0.002	31.980	32.250	32.530
μ_{12}	32.120	0.178	0.002	31.780	32.120	32.480
μ_{22}	33.540	0.168	0.002	33.200	33.540	33.870
μ_{13}	33.520	0.138	0.002	33.250	33.520	33.800
μ_{23}	33.520	0.138	0.002	33.250	33.520	33.800
σ_b^2	0.059	0.181	0.003	0.001	0.021	0.362

0.059; μ_{12} is the mean effect due to fuel B on day 2; μ_{21} is the effect due to fuel T on day 1 and so on. None of the credible intervals contain zero, which implies that all the effects considered are important. However, the lower limit of 95% credible interval of σ_b^2 is very close to zero. As the mean is larger than the median of the distribution of σ_b^2, it implies that the distribution of σ_b^2 is positively skewed which will also be evident in the portrait of kernel density of σ_b^2 presented in Section 2.6.1.

2.5.1.3 Contrast: B2-B1

In Table 2.2 on days 1-3, we treat the first nine tests (day 1, day 2 am) as fuel B1 and the last nine tests (day 2 pm, day 3) as fuel B2 (although we know that both B1 and B2 are actually B). We want to compare the performance of B2 and B1 and ultimately want to study the benefit of switching whether B to B or B to T is better in Section 2.5.1.4.

Mixed Model

We consider the mixed linear model regarding contrast B2-B1.

$$Y_{jkm} = \alpha + \beta_j + \delta_k + \epsilon_{jkm}, \tag{2.10}$$

where Y_{jkm} is the response corresponding to the mth test (m = 1 or 2) on day k (k = 1, 2, 3) corresponding to the test fuel fuel j (j = 1 or 2), α is the intercept, β_j is the effect due to jth fuel, δ_k is the random effect due to kth day, ϵ_{jkm} is the

error term corresponding to the mth test ($m = 1$ or 2) of fuel j ($j = 1$ or 2) on day k ($k = 1, 2, 3$).

Priors and Results

We assume priors for the parameters in model (2.10).

$$\alpha \sim N(0, 0.001), \quad \beta_2 \sim N(0, 0.001)$$
$$\rho \sim \text{beta}(1, 1), \quad \log(\sigma) \sim U(-20, 20).$$

The priors for α , β_2, ρ (intra-class correlation), and $\log(\sigma)$ are assumed to be non-informative. Previously, the prior for α was considered as $\alpha \sim N(38, 0.1)$. As there is no substantial differences in the results either assuming $\alpha \sim N(38, 0.1)$ or $\alpha \sim N(0, 0.001)$ i.e. non-informative prior for α, we use only non-informative priors for α in the subsequent analysis. The results concerning the contrast B2-B1 are presented in Table 2.6. The programme related to this table is given in the Appendix.

Table 2.6: Results for contrast B2-B1

Effect	Mean	SD	MC Error	P(2.5)	Median	P(97.5)
B1	32.140	0.311	0.003	31.600	32.120	32.770
B2	31.730	0.289	0.003	31.190	31.730	32.330
Fuel Diff β_2	-0.411	0.361	0.004	-1.118	-0.403	0.257
σ_b^2	0.115	0.454	0.005	0.005	0.047	0.562

2.5.1.4 Contrast: (T-B)-(B2-B1)

We want to see whether there is a benefit of changing fuels in the car i.e. switching B to T or B to B has any extra benefit or not. Data regarding contrast (T-B)-(B2-B1) is given in Table 2.7.

Mixed Model

We have created three dummy variables to present the mixed model (2.11).

Table 2.7: Data for contrast(T-B)-(B2-B1)

Day	Fuel	Y
1	B1	31.8666
1	B1	32.4208
2	B1	31.9774
2	B2	31.5772
3	B2	31.6541
3	B2	31.9365
4	B	32.3205
4	B	32.2858
5	B	31.9791
5	T	33.6102
6	T	33.3391
6	T	33.6334

$$Y_{jk} = \alpha + \beta_2 D_{2k} + \beta_3 D_{3k} + \beta_4 D_{4k} + \delta_k + \epsilon_{jk} \tag{2.11}$$

where Y_{jk} is the response corresponding to the test fuel j (j = 2, 3, 4) on day k (k = 1, 2, ..., 6), α is the intercept, β_j is the effect due to the jth fuel (note that j= 2, 3, and 4 corresponds to fuel 'B2', 'B', and 'T' respectively), D_{2k} is the dummy variable corresponding to fuel B2, D_{3k} is the dummy variable corresponding to fuel B, D_{4k} is the dummy variable corresponding to fuel T, δ_k is the random effect due to kth day, ϵ_{jk} is the error term corresponding to the test fuel j (j = 2, 3, 4) on day k (k = 1, 2, ..., 6).

Thus, we have the mean effects α, $\alpha+\beta_2$, $\alpha+\beta_3$, and $\alpha+\beta_4$ corresponding to fuels B1, B2, B and T respectively.

Priors and Results

We assume priors for the parameters in model (2.11):

$$\alpha \sim N(0, 0.0001), \quad \beta_j \sim N(0, 0.0001), \quad j = 2, 3, 4;$$

$$\rho \sim \text{beta}(1, 1), \quad \log(\sigma) \sim U(-20, 20).$$

Table 2.8: Results from contrast (T-B)-(B2-B1)

Effect	Mean	SD	MC Error	P 2.5	Median	P 97.5
α (B1)	32.090	0.187	0.004	31.710	32.090	32.460
β_2	-0.365	0.222	0.004	-0.802	-0.371	0.080
β_3	1.411	0.232	0.006	0.968	1.409	1.912
β_4	1.776	0.320	0.008	1.138	1.778	2.424
B2	31.720	0.173	0.003	31.380	31.720	32.070
B	32.150	0.179	0.004	31.790	32.160	32.500
T	33.560	0.180	0.005	33.210	33.560	33.940
B2-B1	-0.365	0.222	0.004	-0.802	-0.371	0.080
T-B	1.411	0.232	0.006	0.968	1.409	1.912
(T-B)-(B2-B1)	1.776	0.320	0.008	1.138	1.778	2.424
σ_b^2	0.033	0.052	0.001	0.001	0.017	0.177

From Table 2.8 we see that the difference between B2 and B1 might be zero as the 95% credible interval for B2-B1 (-0.802, 0.080) includes zero. However, the credible interval for (T-B)-(B2-B1) does not include zero which implies that there might be a benefit of switching B to T rather than switching B1 to B2.

2.5.1.5 Nested Models

In this stage we will look at "fuel within week" as a fixed effect and "day within week" as a random effect and the corresponding data were presented in Table 2.3.

$$Y_{ijkm} = \alpha + \omega_i + \phi_{ij} + \delta_{ik} + \epsilon_{ijkm} \tag{2.12}$$

where Y_{ijkm} is the response corresponding to the mth test ($m = 1$ or 2) of the jth fuel ($j = 1$ or 2) on the kth day ($k = 1, 2, 3$) in the ith week, ($i = 1, 2$). α is the mean fuel effect/ intercept, ω_i is the effect due to week i, ϕ_{ij} is the effect due to fuel j in week i, and δ_{ik} is the random effect due to kth day in week i. Also note that m=1 or 2 corresponds to 'am' or 'pm' and j=1 or 2 corresponds to fuel 'B' or 'T' respectively in the model. For the model (2.12) we impose two corner constraints for two set of parameters: ω_1=0 is for identifying effect of week (we consider week 1 as reference category) and ϕ_{i1}=0, $i = 1, 2$ is for identifying effect of a fuel in i-th week, where base fuel (fuel 1) is reference category.

Priors

We assume the following priors for the model (2.12).

$$\alpha \sim N(0, 0.001), \quad \omega_2 \sim N(0, 0.001), \quad \phi_{i2} \sim N(0, 0.001)$$

$$\rho \sim \text{beta}(1, 1), \quad \log(\sigma) \sim U(-20, 20)$$

The prior for α has been set to be normal by specifying the mean centered at 0 and a low precision i.e. a non-informative prior has been considered for α with mean 0 and variance 1000 . The priors for ω_2, ρ, and $\log(\sigma)$ are also specified as non-informative. The results concerning the contrast B2-B1 are presented in Table 2.9. The programme related to this table is given in the Appendix.

Results

We present the results concerning model (2.12) in Table 2.9. The fixed effect due to

Table 2.9: Results from the nested model

Effect	Mean	SD	MC Error	P(2.5)	Median	P(97.5)
α	32.230	0.202	0.004	31.840	32.230	32.630
ω_2	-0.106	0.285	0.006	-0.663	-0.107	0.435
ϕ_{12}	-0.429	0.238	0.005	-0.910	-0.424	0.047
ϕ_{22}	1.542	0.241	0.005	1.084	1.546	1.975
σ_b^2	0.045	0.078	0.002	0.005	0.027	0.181

week 2 (ω_2) might have no effect on the response variable as 95% credible interval of ω_2 includes zero. Also, the effect of fuel T (ϕ_{22}) on week 2 has positive effect on the response though week 1 does not seem to have similar effect as ϕ_{12} contains zero in its 95% credible interval.

2.5.2 Round Robin Experiments

A round robin is a test programme in which a number of laboratories test identical samples of a number of test materials primarily in order to determine the precision (repeatability and reproducibility) of a test method.

Round robin programmes are carried out to understand and quantify the variation in a test method. They may also be used to measure the performances of particular fluids or fluid batches and/or to investigate the strength of a test method.

Model

Let us consider, for example, a linear mixed model to interpret a round robin analysis.

$$Y_{ijk} = \alpha + \delta_i + \epsilon_{ijk} \tag{2.13}$$

where Y_{ijk} is the kth repeat on fuel j at lab i and $Y_{ijk}|\mu_{ij} \sim \mathrm{N}(\mu_{ij}, \tau)$ with $\mu_{ij} = \alpha + \delta_i$ which is the mean for ith lab corresponding to jth fuel, δ_i is the lab-to-lab error term with mean zero and standard deviation σ_L, $\tau = 1/\sigma^2$ and ϵ_{ijk} is the within-lab error term with mean zero and standard deviation σ.

Repeatability and Reproducibility

The terms repeatability and reproducibility are defined mathematically as

$$\mathrm{r} = 2.8 \times \sigma \quad \text{and} \quad \mathrm{R} = 2.8 \times \sqrt{(\sigma_L^2 + \sigma^2)} \tag{2.14}$$

respectively. In words we can describe these terms as follows.

Repeatability (r): If two tests are conducted on the same sample at the same lab, then we can be 95% confident they will differ by less than r.

Reproducibility (R): If two tests are conducted on the same sample at different labs, then we can be 95% confident they will differ by less than R.

Example Data Set

Fuels **A** and **B** are tested at a number of labs in order to determine the precision of the test method. Data for round robin analysis of fuels A and B are given in Table 2.10. Both of the fuels are analyzed separately and, for each fuel, the objective is to estimate the mean, repeatability r and reproducibility R.

Table 2.10: Round Robin data

Lab	A	B
1	22.320	49.630
2	23.060	49.360
3	18.440	48.430
4	14.850	51.120
5	22.950	51.540
7		50.290
7		59.470
8		52.600
9	17.530	
9	23.740	
10	18.380	
10	23.490	
11	25.740	57.280
11	24.430	40.530
13	18.620	44.350
13	29.040	48.350
14	24.570	51.640
14	18.230	49.900
16		50.400

2.5.2.1 Round Robin Analysis for Fuel A

The general mixed model for round robin is given in equation (2.13). We assume priors for the parameters relevant to model (2.13) as follows

$$\alpha \sim N(0, 0.0001), \quad \rho \sim \text{beta}(1, 1), \quad \log(\sigma) \sim U(-20, 20)$$

We assume non-informative priors for ρ, α and $\log(\sigma)$ and the prior δ_i is derived from available information. The results of round robin analysis on fuel A are given in Table 2.11. The mean effect of fuel A is 21.610 and the repeatability is 10.660 which implies that if fuel A is tested in the same lab, then we are 95% confident that tests will differ by less than 10.660. Further the reproducibility for the same fuel is 12.190 which implies that if fuel A is tested in different labs we are 95% sure that they will differ by less than 12.190.

Table 2.11: Round Robin results of fuel A

Effect	Mean	SD	MC Error	P(2.5)	Median	P(97.5)
α	21.610	1.246	0.011	19.080	21.600	24.070
r	10.660	2.251	0.020	7.171	10.320	15.970
R	12.190	2.734	0.028	8.206	11.730	18.880
σ^2	15.130	6.852	0.064	6.560	13.580	32.550
σ_L^2	4.773	6.013	0.073	0.108	2.910	20.860

Table 2.12: Round Robin results of fuel B

Effect	Mean	SD	MC Error	P(2.5)	Median	P(97.5)
α	50.310	1.385	0.014	47.520	50.320	53.060
r	12.720	2.665	0.027	8.709	12.310	19.030
R	14.140	3.110	0.031	9.592	13.650	21.620
σ^2	21.530	9.734	0.099	9.675	19.320	46.190
σ_L^2	5.221	6.372	0.069	0.127	3.271	22.350

2.5.2.2 Round Robin Analysis for Fuel B

For the round robin analysis of fuel B we use model (2.13) and the priors, the same as for fuel A, are given in Section 2.5.2.1. Round robin results obtained for fuel B is presented in Table 2.12.

The mean effect of fuel B is 50.310 shown in Table 2.12. The repeatability is 12.720 which implies that if fuel B is tested in the same lab, then we are 95% confident that tests will differ by less than 12.720. Further, the reproducibility for the same fuel is 14.140 which implies that if fuel B is tested in different labs we can be 95% sure that they will differ by less than 14.140.

2.6 Convergence and MCMC

Convergence is an important issue for the correct estimation of the posterior distributions of interest. Convergence refers to the situation when the Gibbs sampler or other MCMC techniques eventually reach a stationary distribution. We need to identify at which point convergence takes place and to determine the burn-in period. It is also necessary to determine the number of iterations that would be enough to summarize posterior distributions after reaching stationary distributions. We have monitored convergence diagnostics for the parameters of the models in

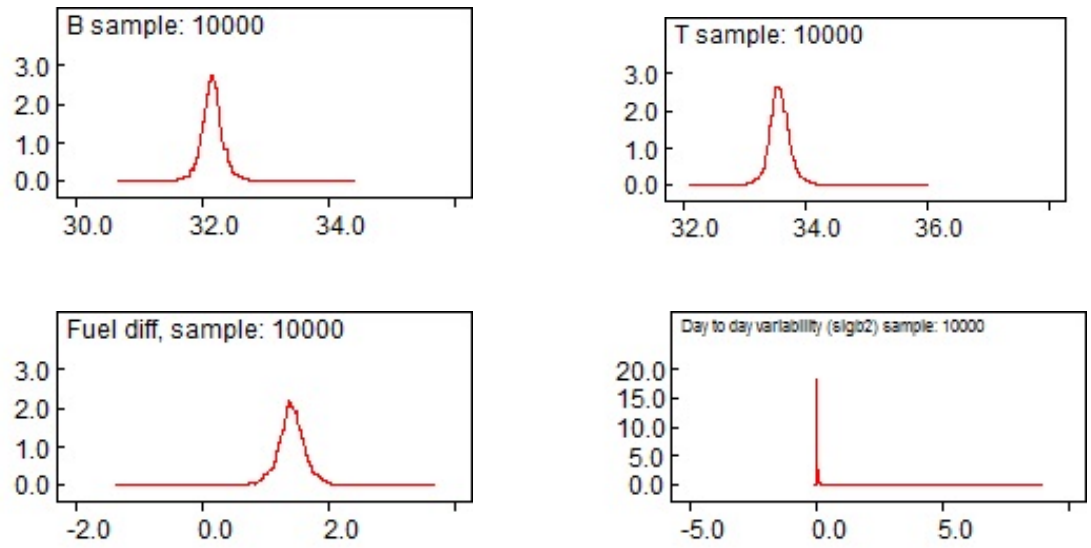

Figure 2.1: Kernel densities of parameters in fuel economy experiment; Base fuel (B) (top left), Test fuel (T) (top right), Fuel Difference β_2 (bottom left), Day to day variability (σ_b^2) (bottom right).

the fuel economy and round robin experiments.

2.6.1 Convergence Diagnostics for Fuel Economy Experiments

There are several ways to monitor convergence. We begin with the Monte Carlo error (MC error) which measures the variability of each estimate due to the simulation. MC error must be low in order to calculate the parameter of interest with increased precision. A rule of thumb is that MC errors should be less than 5% of the corresponding posterior standard deviations (SD). In Table 2.5 all of MC errors of posteriors were less than 5% of the respective standard deviations.

A graphical presentation of the posterior density can be obtained by using a technique known as kernel density estimation. The idea is similar to drawing histograms. Often non-convergence is reflected in multimodal distributions. In that case the kernel density is not only multi-modal but also lumpy. We find in Figure 2.1 that kernel densities of posteriors corresponding to the parameters α, β_2, test fuel and σ_b^2 are approximately smooth and unimodal. There is no visible non-

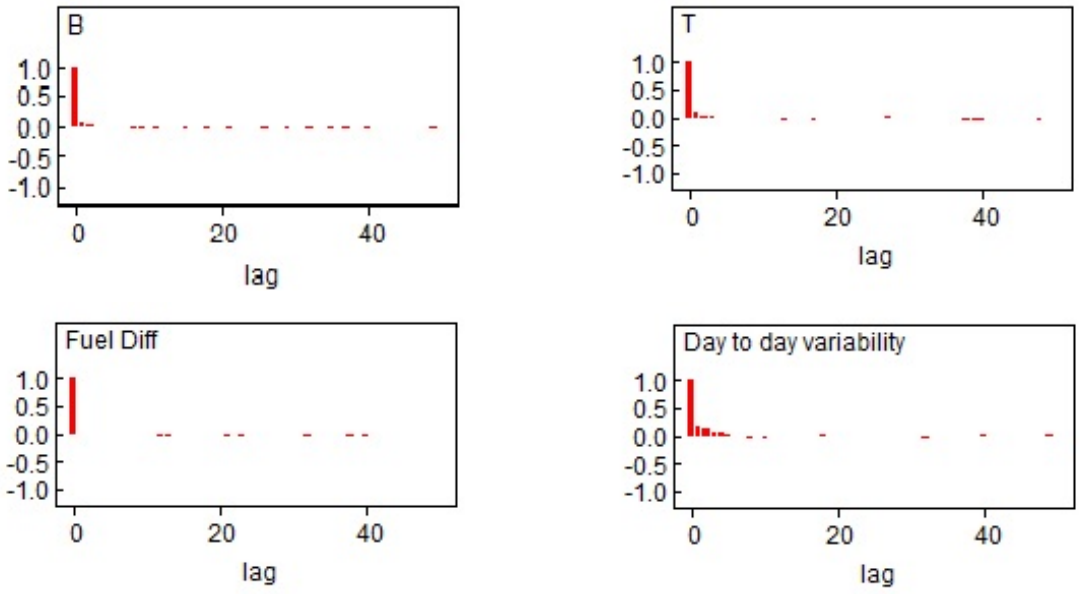

Figure 2.2: Autocorrelation of parameters in fuel economy experiment; Base fuel (B) (top left), Test fuel (T) (top right), Fuel Difference β_2 (bottom left), Day to day variability (σ_b^2) (bottom right).

convergence problem in the kernel densities. If they would have lumpy posteriors, we would let the algorithm run a bit longer to obtain a more reasonable summary of the posteriors.

Autocorrelation refers to a pattern of serial correlation in the chain, where sequential draws of parameter from the conditional distribution are correlated. The autocorrelation happens, because any MCMC method is a Markov chain, therefore, every value in a sample is dependent on the previous values. Because of the autocorrelation, the Gibbs sampler will explore the entire posterior distribution slowly. Generally the level of autocorrelation declines with the increasing lags between sample points in the chain. For example, if we do sub-sampling of every tenth iterate (lag 10), the level of autocorrelation often declines. If this dampening does not happen then we should think that there is a problem and it would be necessary to re-parameterize the model. There was some evidence of sampling autocorrelation in the parameters before thinning but autocorrelation effects disappeared when we thinned the data by taking every 15th point and these are reflected in Figure 2.2.

Trace/history plot which is an intuitive and easily implemented convergence diag-

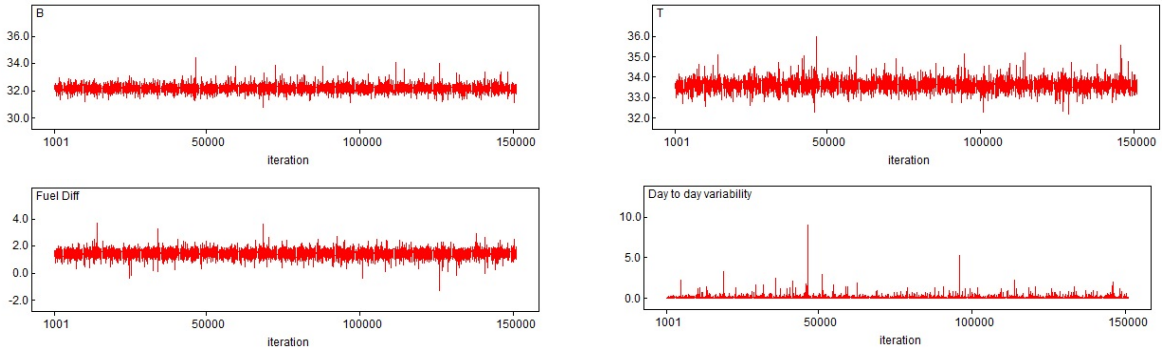

Figure 2.3: History plot of parameters in fuel economy experiment; Base fuel (B) (top left), Test fuel (T) (top right), Fuel Difference β_2 (bottom left), Day to day variability (σ_b^2) (bottom right).

nostic tool, plots the parameter value at time t against the iteration number. If a model is converged the history plot will move around the mode of the distribution. If there is some trending in the sample space, this will be a clear sign of non-convergence. If multiple chains are run, then all chains are visually separated in the graph and there will be well mixing of chains in the event of reaching convergence. In Figure 2.3 we find the well mixing of two chains which provides evidence of convergence.

The Gelman-Rubin (GR) statistic [Gelman and Rubin, 1992] is available as a convergence diagnostic, when two or more chains are generated in parallel, each one starting from different initial values. This is an ANOVA-type diagnostic test which is implemented by calculating and comparing the between-sample and the within-sample variability (i.e., intersample and intrasample variability). The statistic GR can be estimated by

$$GR = \frac{\hat{V}}{\text{WSV}} = \frac{\acute{S}-1}{\acute{S}} + \frac{\text{BSV}/\acute{S}}{\text{WSV}}\frac{\nu+1}{\nu}$$

where ν is the number of generated samples/chains, $\acute{S}$ is the number of iterations kept in each sample/chain, $\text{BSV}/\acute{S}$ is the variance of the posterior mean values over all generated samples/chains (between-sample variance), WSV is the mean of the variances within each sample (within-sample variability), and

$$\hat{V} = \frac{\acute{S}-1}{\acute{S}}\text{WSV} + \frac{\text{BSV}}{\acute{S}}\frac{\nu+1}{\nu}$$

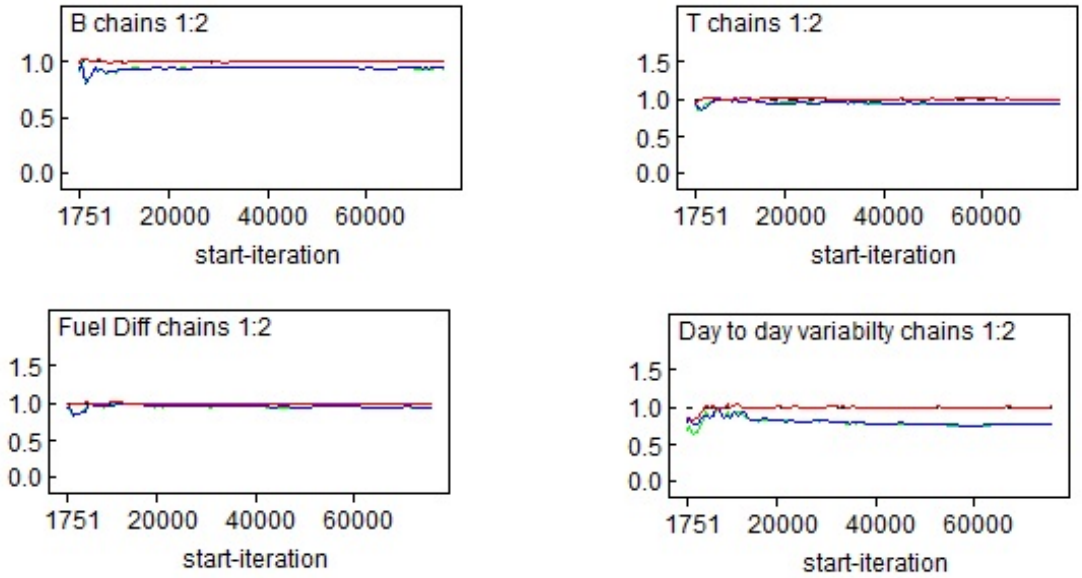

Figure 2.4: Gelman Rubin statistics of parameters in fuel economy experiment; Base fuel (B) (top left), Test fuel (T) (top right), Fuel Difference β_2 (bottom left), Day to day variability (σ_b^2) (bottom right).

is the pooled posterior variance estimate. When convergence is achieved and the size of the generated data is large, then Gelman-Rubin (GR) statistic should approximately be equal to 1. The drawback of this statistic is that its value depends greatly on the choice of initial values.

Figure 2.4 displays the examples of Gelman-Rubin statistics (GR). We started with two sets of initial values for all parameters. It seems that the GR statistics are approximately equal to 1 for all the parameters except for the parameter due to day to day variability (σ_b^2). Also, an unusual autocorrelation pattern was observed with regard to σ_b^2 in Figure 2.2 which might be an indication of less accuracy of σ_b^2 estimation. To sum up the idea we conclude that convergence is achieved for most of the parameters though we are doubtful about σ_b^2. However, simulation studies would be useful to assess the accuracy of all the parameter estimation including σ_b^2.

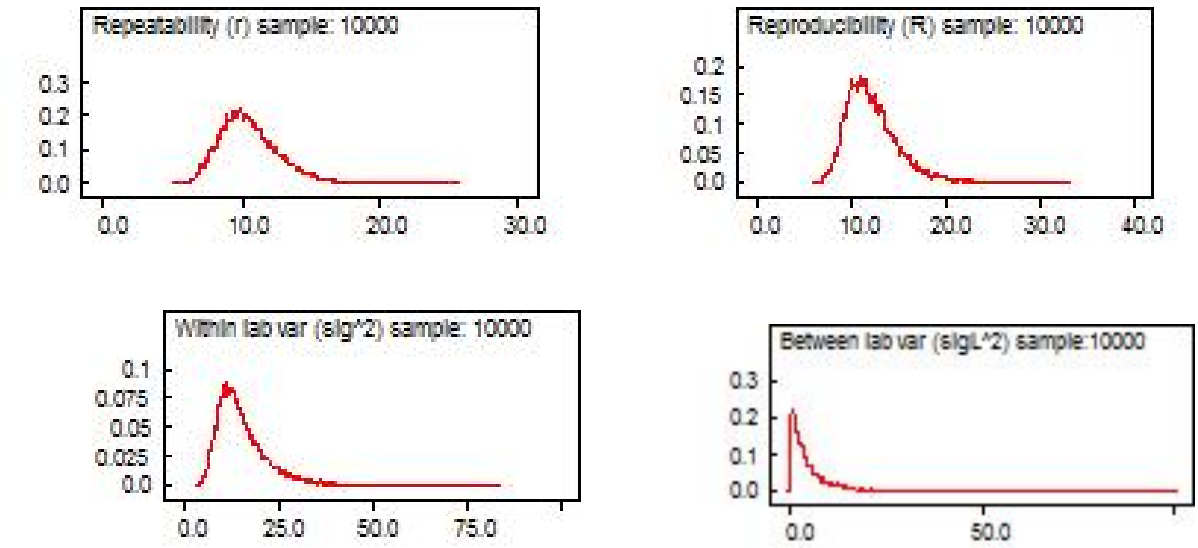

Figure 2.5: Kernel densities of parameters in round robin experiments; Repeatability (r) (top left), Reproducibility (R) (top right), Within lab variability (σ^2) (bottom left), Between lab variability (σ_L^2) (bottom right).

2.6.2 Convergence Diagnostics for Round Robin Experiments

First, we examine the Monte Carlo (MC) errors for the parameters of the linear mixed model (2.13) in the round robin experiment. In Table 2.11 all of MC errors of posteriors were less than 5% of the respective standard deviations.

We find in Figure 2.5 that kernel densities of posteriors corresponding to the parameters r, R, σ^2, σ_L^2 are approximately smooth and unimodal. There is no visible non-convergence problem in the kernel densities.

There was some evidence of sampling autocorrelation in the parameters before thinning (see Figure 2.6), but autocorrelation effects disappeared when we thinned the data by taking every 20th point and these are reflected in Figure 2.7. In Figure 2.8 the history plots show that there is good mixing of two chains which provides evidence of convergence.

The Figure 2.9 displays examples of Gelman-Rubin statistics (GR). We started

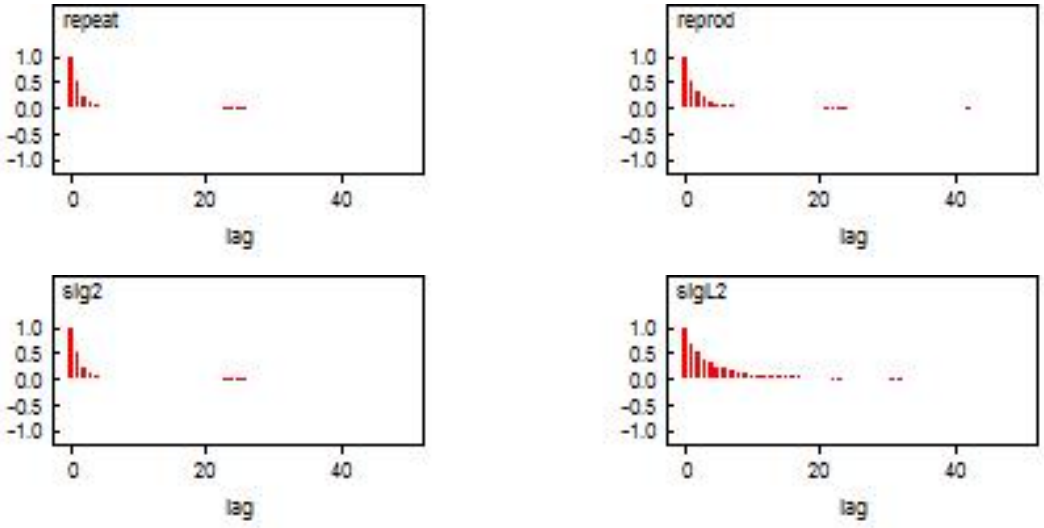

Figure 2.6: Autocorrelation of parameters (before thinning) in round robin experiments; Repeatability (r) (top left), Reproducibility (R) (top right), Within lab variability (σ^2) (bottom left), Between lab variability (σ_L^2) (bottom right).

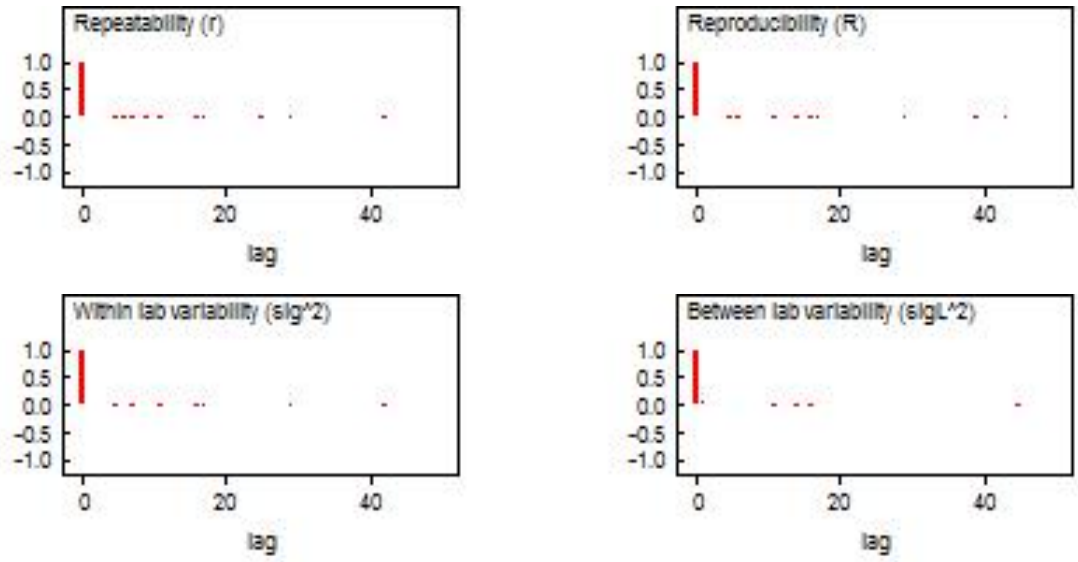

Figure 2.7: Autocorrelation of parameters (after thinning) in round robin experiments; Repeatability (r) (top left), Reproducibility (R) (top right), Within lab variability (σ^2) (bottom left), Between lab variability (σ_L^2) (bottom right).

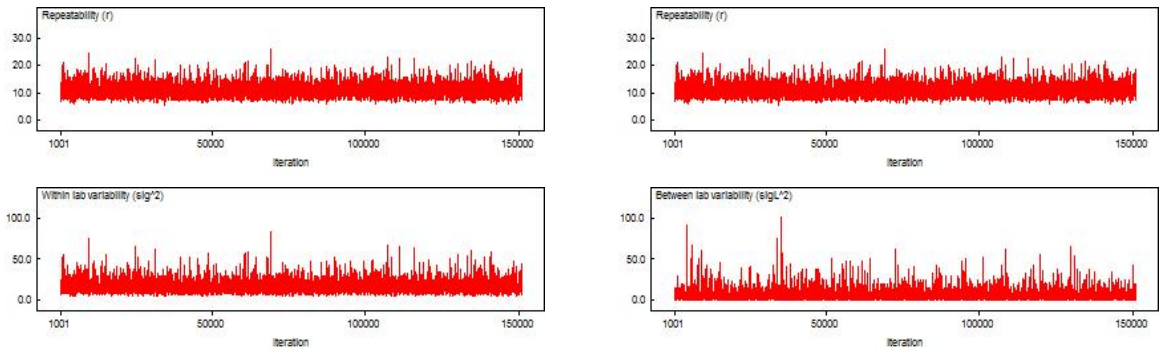

Figure 2.8: History plot of parameters in round robin experiments; Repeatability (r) (top left), Reproducibility (R) (top right), Within lab variability (σ^2) (bottom left), Between lab variability (σ_L^2) (bottom right).

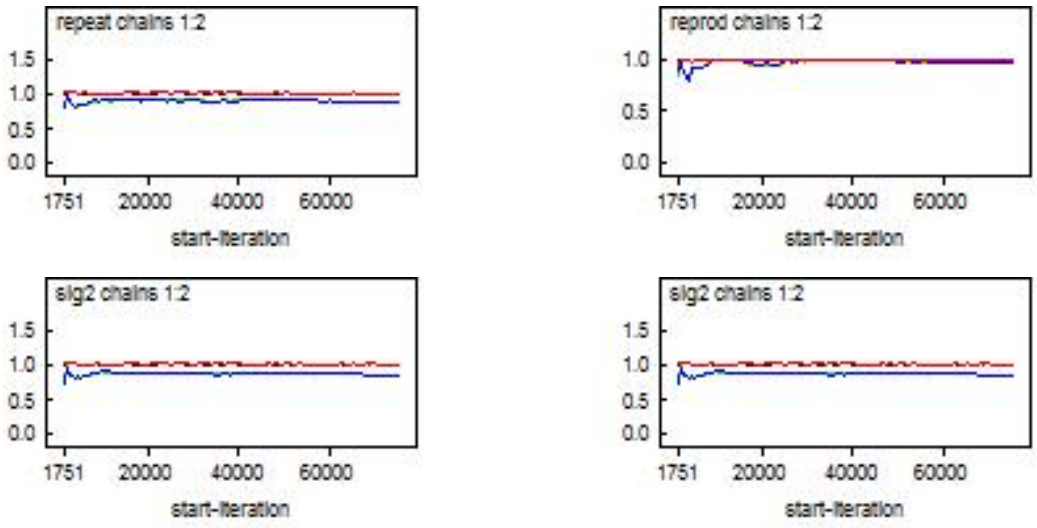

Figure 2.9: Gelman Rubin (GR) statistics of parameters in round robin experiments; Repeatability (r) (top left), Reproducibility (R) (top right), Within lab variability (σ^2) (bottom left), Between lab variability (σ_L^2) (bottom right).

Table 2.13: Robustness of estimates with different priors

Priors		Posterior Estimates				
Between Lab	Within Lab	$\hat{\alpha}$	$\hat{\sigma^2}$	$\hat{\sigma_L^2}$	Repeatability **(r)**	Reproducibility **(R)**
$\rho \sim$ Beta(1, 1)	log(σ) ~U(-20, 20)	21.61	15.13	4.773	10.66	12.19
$\rho \sim$ Beta(1.5, 1.5)	log(σ) ~U(-20, 20)	21.59	14.52	5.780	10.43	12.31
$\rho \sim$ Beta(2.5, 2.5)	log(σ) ~U(-20, 20)	21.54	14.05	7.355	10.27	12.64
$\sigma_L \sim$ U(0, 100)	log(σ) ~U(-20, 20)	21.61	16.97	4.404	11.26	12.58
$\tau_L \sim$ Gamma(0.1, 0.1)	log(σ) ~U(-20, 20)	21.63	16.87	1.879	11.24	11.86

with two sets of initial values for all parameters. It seems that the GR statistics are approximately equal to 1 for all the parameters. Therefore, we may conclude that there is no evidence against the convergence for all parameters.

In our study we find that the parameters of the round robin model for fuel A passed the convergence criteria. However, the convergence diagnostics work like "alarms" that detect certain problems concerning the convergence of the chain. Since the focus of each diagnostic is different, to ensure convergence all tests must be passed, not rejected [Ntzoufras, 2009].

2.7 Robustness of Posterior Distributions in Round Robin Experiments

Robustness of the posterior distribution is an important issue in Bayesian modelling. We can assess how robust the posterior distributions are to the selection of the prior distributions via sensitivity analysis. When prior information is available, sensitivity analysis focuses on the structure of the prior distribution; when non-informative priors are used, it focuses on how different choices of prior parameters (prior mean or variance) influence the posterior inference.

We consider round robin analysis of fuel A as a test example to see how choice of prior parameters influences the posterior estimates. We have monitored the robustness of posterior distributions assuming different values of the prior mean and variance and also assuming different families of priors as shown in Table 2.13.

Different priors have negligible impact on the posterior estimates, specifically posterior estimates of intercept (α) is unaffected by the choice of different priors. Esti-

mates of repeatability (r) and reproducibility (R) slightly varying when we consider slightly informative gamma priors for the precision of between lab variability. In this case, though the estimate of within lab variability ($\hat{\sigma}^2$) is uninfluenced but between lab variability ($\hat{\sigma}_L^2$) is greatly influenced by the chosen informative gamma prior. Thus, the posterior estimates of the intercept, within lab variability, repeatability and reproducibility are not varying too much with different choices of priors. However, robustness of posterior distributions will be explored more in simulation studies given in Section 2.10.

2.8 Likelihood Methods in Fuel Economy Experiments

The objective of this chapter was to analyze fuel economy experimental data by Bayesian methods as likelihood-based methods were not enough to deal with the estimation of variance component parameters. However, to have some idea about likelihood-based methods we have presented analyses by giving two examples, particularly by likelihood analysis of contrasts T-B and (T-B)-(B2-B1).

2.8.1 Contrast: T-B

We have fitted linear mixed-effect models for contrast T-B by residual maximum likelihood (REML) and maximum likelihood (ML) methods using the statistical software R. The methods of REML and ML for linear mixed effects are described in Pinheiro and Bates [2000]. The results from likelihood-based methods for contrast T-B are presented in Table 2.14 and Table 2.15.

Table 2.14: Linear Mixed-effects model fit for (T-B) by REML method

Effect	Value	SE	DF	t-value	p-value
α (B)	32.195	0.102	2	316.537	0.000
β_2	1.332	0.144	2	9.264	0.012
$\hat{\sigma}_b$	1.683734×10^{-6}				

The likelihood estimates of fixed effects presented in Tables 2.14 and 2.15 and the corresponding Bayesian estimates in Table 2.5 are similar. For example, mean effects of fuels B and T are almost the same in both methods. However, the Bayesian

Table 2.15: Linear Mixed-effects model fit for (T-B) by ML method

Effect	Value	SE	DF	t-value	p-value
α (B)	32.195	0.102	2	316.537	0.000
β_2	1.332	0.144	2	9.264	0.012
$\hat{\sigma}_b$	1.849246×10^{-6}				

and likelihood estimates of variance components, particularly the variance of random effects due to day differs substantially. The between day standard deviations obtained by REML and maximum likelihood methods are 1.683734×10^{-6} and 1.849246×10^{-6} respectively. It should be noted that by default ML or REML method in R forces the estimate of variance components to be nonzero. Though we obtained the positive estimates of variance components in likelihood based methods but truly it is zero or very close to zero which is unrealistic. However, the corresponding Bayesian estimate of the variance component ($\hat{\sigma}_b^2$) found in Table 2.5 is 0.059. Thus, a poorly estimated variance component in likelihood-based methods becomes estimable in the Bayesian method assuming some priors. However, the quality of the Bayesian credible interval for $\hat{\sigma}_b^2$ will be compared and examined with the profile likelihood and bootstrap based confidence intervals in Section 2.9 and through simulation studies in Section 2.10.

2.8.2 Contrast: (T-B)-(B2-B1)

The likelihood and Bayesian estimates of fixed effects corresponding to contrast (T-B)-(B2-B1) are similar shown in Table 2.16, Table 2.17 and Table 2.8 respectively. The estimates for between days standard deviations ($\hat{\sigma}_b$) are 1.50999×10^{-6} and 2.312336×10^{-6} in REML and ML method respectively, implying that each of the between days variance component is approximately 0 which is unrealistic. The Bayesian method provides a reasonable estimate of the between days variance component ($\hat{\sigma}_b^2$) as 0.035 shown in Table 2.8.

Table 2.16: Linear Mixed-effects model fit for (T-B)-(B2-B1) by REML method

Effect	Value	SE	DF	t-value	p-value
α (B1)	32.088	0.124	5	259.225	0.000
β_2	-0.366	0.175	3	-2.089	0.128
β_3	0.107	0.175	3	0.610	0.585
β_4	1.439	0.175	3	8.222	0.004
$\hat{\sigma}_b$	1.50999×10^{-6}				

Table 2.17: Linear Mixed-effects model fit for (T-B)-(B2-B1) by ML method

Effect	Value	SE	DF	t-value	p-value
α (B1)	32.088	0.124	5	259.225	0.000
β_2	-0.366	0.175	3	-2.089	0.128
β_3	0.107	0.175	3	0.610	0.585
β_4	1.439	0.175	3	8.222	0.004
$\hat{\sigma}_b$	2.312336×10^{-6}				

2.9 Profile Likelihood and Confidence Intervals

The standard method of calculating a confidence interval (CI) is the so-called Wald procedure, which relies on the asymptotic normality of $\hat{\theta}$, the MLE of θ. However, Wald-type intervals can perform badly for small to moderate sample sizes due to poor estimates of sampling variance, bias in the MLE, and/or asymmetry in the sampling distribution of the MLE [Donaldson and Schnabel, 1987; Gimenez et al., 2005]. A construction of a confidence interval that is likely to be more robust in small samples may be derived from the asymptotic χ^2 distribution of the likelihood ratio test statistic. This likelihood-ratio based method is known as 'profile likelihood'. Profile likelihood confidence intervals do not assume normality of the estimates and appear to perform better for small sample sizes than Wald-type confidence intervals. They are, nonetheless, still based on an asymptotic approximation to the chi-square distribution of the log likelihood ratio test statistic.

Let us consider a model with parameters $\beta_1, \beta_2, \ldots, \beta_j, \ldots, \beta_p$. Suppose that $\theta = \beta_j$ is of our parameter of interest and δ, where $\delta = c(\beta_1, \beta_2, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_p)$, is the vector of additional parameter(s). We define the log-likelihood function as $l(\theta, \delta)$. The profile likelihood function for θ is

$$l(\theta) = \max_{\delta} l(\theta, \delta). \tag{2.15}$$

For each value of δ, $l(\theta)$ is the maximum of the likelihood function over the remaining parameters. Therefore, the profile likelihood function is not a real likelihood function; each point on the profile likelihood function is the maximum value of a likelihood function. The idea of a profile likelihood confidence interval is to invert a likelihood-ratio test to obtain a confidence interval for the parameter of interest. The likelihood ratio test statistic of the hypothesis $H_0 : \theta = \theta_0$ equals the difference between $2l$ for the "full" model and $2l$ for the "reduced" model which has θ fixed

at θ_0, i.e. $2\left[l(\hat{\theta},\hat{\delta}) - l(\theta_0,\hat{\delta}_0)\right]$, where $\hat{\theta}$ and $\hat{\delta}$ are MLEs for the full model and $\hat{\delta}_0$ is the MLE of δ for the reduced model with $\theta = \theta_0$. A $100(1-\alpha)\%$ confidence interval for θ is the set of all values θ_0 such that two-sided test of null hypothesis $H_0 : \theta = \theta_0$ would not be rejected at the α level of significance if and only if

$$2\left[l(\hat{\theta},\hat{\delta}) - l(\theta_0,\hat{\delta}_0)\right] < \chi^2_{1-\alpha}(1) \tag{2.16}$$

where $\chi^2_{1-\alpha}(1)$ is the $1-\alpha$ quantile of a χ^2 distribution with 1 degrees of freedom. Further details on profile likelihood can be found in Venzon and Moolgavkar [1988] and Royston [2007].

The statistical package *lme4* in R has impressive facility to provide profile likelihood confidence intervals of parameters under generalized linear models including models with normal responses. This package also enables bootstrap based confidence intervals for the same parameters and thereby facilitates more comparison with Wald-type confidence intervals. Bootstrap methods are computer-intensive method of statistical analysis that use simulation to calculate standard errors, confidence intervals, and significance tests [Davison, 1997; Efron and Tibshirani, 1994]. However, it should be noted that *lme4* does not have the option to provide Wald-type confidence intervals for variance components as the sampling distribution of a variance component estimator is skewed, whereas the basis of Wald-type confidence intervals is asymptotic normality.

We have implemented Wald, profile likelihood and bootstrap based methods for computing confidence intervals for the parameters under model (2.9). Then the likelihood based confidence intervals have been compared with the Bayesian credible intervals shown in Table 2.18 (note that the Bayesian estimates in Table 2.18 are originally from Table 2.5). The point and interval estimates of base fuel are similar in likelihood and Bayesian methods. The Bayesian estimate of fuel difference β_2 is 1.421 which differs from classical counterpart. Also, the Bayesian credible interval for β_2 is slightly conservative (wider) in comparison to profile likelihood or bootstrap confidence intervals. The main idea behind the implementation of profile likelihood, bootstrap based methods, along with the Bayesian technique is to see the differences in point and interval estimates of the variance component, σ_b^2. The point estimate of the day to day variance component (σ_b^2) is zero in classical method which is assumed unrealistic, whereas the Bayesian estimate is nonzero

(0.059). The profile and bootstrap methods show that lower limit of 95% confidence intervals is zero, whereas lower limit of 95% Bayesian credible interval shows that it is nonzero, but close to zero. However, Lambert et al. [2005] notice that if the variance parameter is close to the boundary at zero, MCMC results tend to produce upwardly biased estimates of variance parameters while inferences are based on the posterior mean. Therefore, considering all figures either point or interval estimates we conclude that day to day variance component (σ_b^2) might be nonzero but in between likelihood and Bayesian point estimates in the fuel economy experiment. Yet, a simulation study will examine the performance of classical and Bayesian methods in estimating the variance component (σ_b^2) in Section 2.10.3.

Table 2.18: Likelihood and Bayesian estimates with 95% confidence/credible intervals under different methods in fuel economy experiment

Parameter	Likelihood Method				Bayesian Method	
		95% CI				95% CI
	Estimate	Wald	Profile	Bootstrap	Estimate	
α	32.195	(32.032, 32.358)	(31.915, 32.388)	(32.019, 32.365)	32.170	(31.780, 32.580)
β_2	1.332	(1.102, 1.563)	(1.060, 1.713)	(1.067, 1.598)	1.421	(0.952, 1.877)
σ_b^2	0.000	-	(0.000, 0.099)	(0.000, 0.026)	0.059	(0.001, 0.362)

In round robin real experiment, fuel A was administered once in each of the first five laboratories. To implement linear mixed effect analysis, we need at least two replicates of fuel A in each of the laboratories. So we exclude first five laboratories in the likelihood analysis of the mixed model as well as in profile likelihood and bootstrap based analysis. In Bayesian analysis priors have been assumed as before given in Section 2.5.2.1 and Bayesian results in Table 2.19 will not be same as in Table 2.11 as the number of laboratories are not same in two situations. Table 2.19 shows that the fixed effect estimates are approximately similar though the Bayesian interval estimates are little wider than the classical based intervals. The MLE of σ_L^2 is reported as 0 with a profile likelihood based CI of (0.000, 13.587) and a bootstrap based CI of (0.000, 16.064). The point estimate of σ_L^2 is 0 which is unrealistic, whereas the Bayesian point estimate is reported as 11.517 with 95% credible interval (1.099, 50.313). Therefore, it is seen that there are substantial differences between the likelihood and Bayesian estimates of variance component due to lab (σ_L^2) both in point and interval estimates. Though the Bayesian point estimate seems inflated but note that it lies still within the profile likelihood and bootstrap based confidence intervals. The wider credible interval in the Bayesian analysis is perhaps due to using non-informative priors for σ_L^2 in this small sample

study. However, non-informative priors should not be used when the sample size is low, as suggested by Galindo-Garre et al. [2004], which will be verified in simulation studies also.

Table 2.19: Likelihood and Bayesian estimates with 95% confidence/credible intervals under different methods in round robin experiment for fuel A

Parameter	Likelihood Method				Bayesian Method	
		95% CI				95% CI
	Estimate	Wald	Profile	Bootstrap	Estimate	
α	22.376	(19.944, 24.807)	(19.829, 24.922)	(19.766, 24.865)	22.339	(18.329, 26.073)
σ^2	15.39	-	(6.445, 38.626)	(2.283, 29.866)	17.307	(6.359, 43.770)
σ_L^2	0.000	-	(0.000, 13.587)	(0.000, 16.064)	11.517	(1.099, 50.313)

To conclude this section we summarize that the point estimates of fixed effects do not differ substantially, but Bayesian intervals are slightly wider than corresponding likelihood intervals. The estimate of the variance component might not be zero, but could lie between 0 and the Bayesian point estimate 11.517. The comparatively larger width of the Bayesian 95% credible intervals could be minimized with appropriate choice of priors as noninformative priors might entail wider credible intervals.

2.10 Simulation Studies

As Bayesian methods are commonly used in applied research, more investigation is essential to verify whether Bayesian estimates have better properties than ML estimates and also whether some priors enable better estimates than others [Galindo-Garre et al., 2004]. This investigation and verification can be carried out through simulation studies. These techniques provide empirical estimation of the sampling distribution of the parameters of interest, that cannot be achieved from a single study [Burton et al., 2006].

2.10.1 Performance Measures in Simulation Studies

It is essential to consider the criteria for measuring the performance of the results obtained from different methods or scenarios in simulation studies. The commonly used performance measures are bias, accuracy, and coverage of the parameter estimates. However, Collins et al. [2001] suggested examining more than one per-

formance measure such as mean squared error (MSE) and width of the confidence intervals (e.g. average and median width of 95% confidence/credible intervals) as results may vary across criteria. The main interest in simulation studies is to recognize the sampling distributions of the simulated estimates and therefore the average and variance of all the estimates over all simulations is used to calculate the accuracy measures such as the bias and MSE. A brief description of commonly used performance measures can be provided in the following.

Assessment of Bias

The bias is the deviation of an estimate from the true value which can be obtained by taking the difference of average of an estimate and the true parameter value. For example, the bias of an estimator $\hat{\beta}$ of β is estimated by

$$\text{Bias}(\hat{\beta}) = \sum_{s=1}^{S} \frac{(\hat{\beta}_s - \beta)}{S} = \bar{\hat{\beta}} - \beta \tag{2.17}$$

where $\hat{\beta}_s$ is the estimate of β obtained from the sth simulated data set, $\bar{\hat{\beta}} = \sum_{s=1}^{S} \hat{\beta}_s / S$ is an estimate of the true parameter β and S is the number of accomplished simulations. Another approach of bias assessment is to compute the bias as a percentage of the true value, called percentage relative bias, as

$$\frac{\text{Bias}(\hat{\beta})}{\beta} \times 100. \tag{2.18}$$

As true values of the parameters in simulations might be different in magnitude relative bias can be more representative than the absolute bias during comparison of biases if parameters are on a common scale. One more method of bias assessment is standardized bias which is the bias as a percentage of $\text{SE}(\hat{\beta})$, that is

$$\frac{\text{Bias}(\hat{\beta})}{\text{SE}(\hat{\beta})} \times 100 \tag{2.19}$$

which can be more informative as the consequence of the bias depends on the size of the uncertainty in the parameter estimate. A standardized bias greater than 10% in either direction can have noticeable adverse effects on the efficiency, coverage and error rates [Collins et al., 2001].

Assessment of Accuracy and Precision

Overall accuracy and precision of an estimator can be assessed by mean squared error (MSE), the average squared difference between the estimate and its target. The MSE of $\hat{\beta}$ is estimated by

$$\text{MSE}(\hat{\beta}) = \left(\bar{\hat{\beta}} - \beta\right)^2 + \left(\text{SE}(\hat{\beta})\right)^2, \tag{2.20}$$

where $(\bar{\hat{\beta}} - \beta)$ is the bias, $\text{SE}(\hat{\beta})$ is the empirical SE of the estimate of interest over all simulations. The MSE enables a useful measure of overall accuracy as it takes into account both measures of bias and variability. For easier interpretation, square root of MSE (RMSE) that transforms MSE back onto the same scale as the parameter can be reported [Collins et al., 2001]

Assessment of Coverage

The coverage of a confidence interval is the proportion of times that the obtained confidence interval includes the true parameter value i.e. the proportion of times the $100(1-\alpha)\%$ confidence interval $\hat{\beta}_s \pm Z_{1-\alpha/2}\text{SE}(\hat{\beta}_s)$ includes β, for $s = 1, 2, \ldots, S$. The empirical coverage probability can be obtained from

$$\text{CP}(\hat{\beta}) = \frac{1}{S}\sum_{s=1}^{S} I\left\{|\hat{\beta}_s - \beta| \leq Z_{1-\alpha/2} \times \text{SE}(\hat{\beta}_s)\right\} \tag{2.21}$$

where I{} is an indicator variable. The coverage should approximately be equal to the nominal coverage rate. For example, if the nominal coverage rate is 95%, over-coverage, where coverage rates are above 95%, suggests that the results are too conservative. In practice, so we will not find significant results in many cases when there is a true effect thus leading to a loss of statistical power with too many type II errors. On the other hand, under-coverage, where coverage rates are below 95%, is unacceptable as it indicates over-confidence in the estimates since many real situations will incorrectly detect a significant result, which leads to higher than expected type I errors. A possible criterion for acceptability of the coverage is that the coverage should not fall outside of approximately two SEs of the nominal coverage probability (p), $\text{SE}(p) = \sqrt{p(1-p)/S}$ [Tang et al., 2005]. For example, if 95% confidence intervals are calculated using 2000 independent simulations then $\text{SE}(\hat{p})$ is 0.0048734 and hence between 94.02% and 95.98% of the confidence inter-

vals should include the true parameter value.

In the Bayesian case, the coverage probabilities are calculated on the basis of 95% credible intervals that can be calculated in several ways. Credible intervals are not unique on a posterior distribution. A method for defining a suitable credible interval would be to choose the narrowest interval, which for a unimodal distribution might involve choosing those values of highest probability density including the mode. A credible interval might based on the interval where the probability of being below the interval is as likely as being above it. This interval will include the median. Assuming the mean exists, a credible interval might choose the interval for which the mean is the central point. However, in the context of decision theory, an optimal credible interval will always be a highest probability density set [OHagan, 1994]. For simplicity, we restrict our calculation of credible intervals based on the second definition discussed above i.e. on the basis of the $\alpha/2$-th and $(1-\alpha/2)$-th posterior quantiles. If β_L^*, β_U^* are the $\alpha/2$th and $(1-\alpha/2)$-th posterior quantiles for β respectively, then (β_L^*, β_U^*) is a $100(1-\alpha)\%$ credible interval for β, where $\mathrm{P}[\beta < \beta_L^*] = \alpha/2$ and $\mathrm{P}[\beta > \beta_U^*] = \alpha/2$.

The average width of $100(1-\alpha)\%$ confidence interval is often used as an evaluation criterion in simulation studies, particularly when one method has a similar or higher rate of coverage than another but yields substantially shorter intervals. Shorter intervals translate into greater accuracy and higher power [Collins et al., 2001]. The average $100(1-\alpha)\%$ confidence interval width for a likelihood estimate is calculated as

$$\frac{\sum_{s=1}^{S} 2Z_{1-\alpha/2}\mathrm{SE}(\hat{\beta}_s)}{S}. \tag{2.22}$$

The corresponding average $100(1-\alpha)\%$ credible interval width of a Bayesian estimate is calculated as

$$\frac{\sum_{s=1}^{S} (\beta_{Us}^* - \beta_{Ls}^*)}{S}. \tag{2.23}$$

where β_{Ls}^* and β_{Us}^* are the $\alpha/2$-th and $(1-\alpha/2)$-th posterior quantiles for β respectively at s-th simulation.

2.10.2 Determination of Simulation Size

There are some issues that should be addressed before starting any simulation studies. For instance, the number of simulations to be performed in a study is one of the vital issues. There was no precise justification of choosing the number of simulations in many studies. The number of simulations can vary from 100 to 100000 replications with the most common choices being 1000 and 10000 replications as reported by Burton et al. [2006]. The current practice regarding choice of the number of simulations is based on expected standard error of $\hat{\beta}$, power to detect differences of a specified level of from true value as statistically significant. However, we will adopt the number of simulations on the basis of accuracy of an estimate of interest. The number of simulations (S) can be obtained using the following formula:

$$S = \left(\frac{Z_{1-\alpha/2}\sigma_{\hat{\beta}}}{\delta} \right)^2 \tag{2.24}$$

where δ is the specified level of accuracy of the estimate, that is the acceptable difference from the true value β, $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the standard normal distribution and $\sigma^2_{\hat{\beta}}$ is the variance for the parameter of interest (β).

A realistic estimate of the variance may be obtained from the real data if the simulations are based on a real data set and trying to maintain the same amount of variability. If the variance is unknown or cannot be estimated reliably then it may be possible to perform an identical simulation study to obtain realistic estimates for the variance or consider the measure of accuracy as a percentage of the standard error of the $\hat{\beta}$. The size of simulations to perform is thus dependent on the true value of the estimate of interest, the variability of the estimate of interest, and the required accuracy [Burton et al., 2006].

2.10.3 Simulation Studies on Fuel Economy Experiments

In this section we present the results of simulation studies on fuel economy experiments to observe the performance of point estimators and confidence/credible intervals under likelihood and Bayesian methods.

Table 2.20: Simulated performance of maximum likelihood and Bayesian estimates in fuel economy experiments assuming true parameter values as $\alpha = 32, \beta_2 = 1.4$, $\sigma_b^2 = 0.05$; priors as $\alpha \sim N(0, 0.001)$, $\beta_2 \sim N(0, 0.001)$; and sample size n=6

Parameter	Mean/Posterior Mean			Median/Posterior Median			Coverage Probability	Average width of 95% CI	Median width of 95% CI
	Bias	% Relative Bias	RMSE	Bias	% Relative Bias	RMdSE			
MLE									
α	0.004	0.01	0.217	0.002	0.01	0.144	0.971	1.617	1.514
β_2	-0.005	-0.36	0.284	-0.008	-0.60	0.187	0.929	1.559	1.520
σ_b^2	0.028	56.08	0.103	-0.008	-16.49	0.048	0.979	∞	1.896
$\rho \sim$ Beta(1, 1)									
α	0.059	0.18	0.215	0.037	0.12	0.208	0.957	1.151	1.110
β_2	-0.033	-2.38	0.253	-0.024	-1.70	0.255	0.939	1.242	1.198
σ_b^2	0.112	223.83	0.176	0.006	11.20	0.047	0.991	0.853	0.683
$\rho \sim$ Beta(1.5, 1.5)									
α	0.058	0.18	0.209	0.039	0.12	0.203	0.954	1.112	1.071
β_2	-0.038	-2.72	0.248	-0.028	-1.96	0.248	0.943	1.243	1.224
σ_b^2	0.088	176.89	0.141	0.007	14.36	0.045	0.988	0.719	0.576
$\rho \sim$ Beta(2.5, 2.5)									
α	0.045	0.14	0.205	0.029	0.09	0.201	0.953	1.103	1.060
β_2	-0.032	-2.28	0.247	-0.022	-1.56	0.246	0.948	1.274	1.226
σ_b^2	0.074	148.40	0.121	0.009	17.54	0.045	0.987	0.583	0.466
$\rho \sim$ Beta(3, 3)									
α	-0.004	-0.01	0.203	-0.005	-0.01	0.204	0.953	1.040	1.000
β_2	-0.002	-0.13	0.247	-0.001	-0.06	0.247	0.941	1.244	1.200
σ_b^2	0.062	123.66	0.101	0.009	18.34	0.044	0.980	0.513	0.415
$1/\sigma_b^2 \sim$ Gamma(0.01, 0.001)									
α	0.234	0.731	0.303	0.140	0.438	0.242	1	2.807	2.770
β_2	-0.150	-10.687	0.294	-0.131	-9.391	0.287	1	3.170	3.140
σ_b^2	0.550	1100.205	0.570	-0.013	-26.582	0.017	1	4.406	3.889
$1/\sigma_b^2 \sim$ Gamma(0.1, 0.1)									
α	0.347	1.08	0.392	0.258	0.81	0.318	1.000	2.904	2.772
β_2	-0.200	-14.32	0.311	-0.171	-12.18	0.298	1.000	3.109	3.075
σ_b^2	1.920	3840.10	1.959	0.321	641.96	0.324	0.941	8.702	7.181

The real fuel economy experiment was conducted in three days with two sessions each to compare test and base fuels and thus the sample size was 6. A simulation study has been implemented taking a sample of 6 to realize the differences of results from a single study and from the simulation study.

In simulation studies we generated data from a normal distribution whose parameters are from the model (2.9) by setting arbitrarily the parameters as $\alpha = 32$, β_2=1.4, and σ_b^2=0.05 which might be close to reality as indicated in the Bayesian analysis of real data (see Table 2.5). The number of simulations for the fuel economy experiments was determined using the formula (2.24). We considered three parameters separately to decide about simulation size. Generally the bias in the fixed effects are small if the variation of random effect distribution is small. However, the estimate of the variability of the random effect is always severely biased as observed by Litière et al. [2008]. If we consider fixed effects for calculation of simulation size, it will be very small as bias is very low for the parameter of fixed effects. Therefore, we take into account the random effects during calculation of simulation size and allow bias as 15%. Assuming σ_b^2=0.05 , $Z_{1-\alpha/2}$=1.78, standard deviation of estimate $\sigma_{\hat{\beta}}$=0.175, where $\hat{\beta} = \hat{\sigma_b^2}$, and $\delta = 0.05 \times 0.15$, the simulation size would be 1725. However, also for practical reasons, for example, results are not changing substantially with higher number of simulations (see Table 2.22) and as each set of the runs takes more than 13 hours to implement 5000 simulations, we fix the number of simulations at 2000 which takes approximately 6-7 hours to terminate in the Bayesian analysis. However, for likelihood analysis 2000 simulations take only a few minutes to terminate.

The lme function in the nlme package of R provides facilities to calculate the estimates of all parameters relevant to fixed and random effects including their 95% confidence intervals in likelihood-based methods. In Table 2.20 the first column denotes the parameters, second column presents the bias of the parameter estimates, the third column is for percent relative bias, the fourth column is for root mean squared error (RMSE), the seventh column represents root median squared error (RMdSE) and rest of the columns are self explanatory. In the table, the first set of results (first three rows) is for likelihood estimates, obtained through restricted maximum likelihood (REML) procedures, corresponding to parameters α, β_2, and σ_b^2 computed from 2000 simulations. We see the bias and relative bias based on the

mean are larger than the median based bias and relative bias for all parameters. This happens because mean estimates are affected by unusual (extreme) estimates that arose during simulations. The estimation of the random effects are not of our primary interest, the main concern being to estimate fixed effects, particularly, the difference of the effects of test and base fuels (β_2). The coverage probability is 0.929 which is significantly below the acceptable range, because, by definition, a 95% confidence interval should have coverage probability of at least 0.95. However, even if the true coverage probability equals 95%, the coverage probability obtained from a simulation study might not be exactly equal to 0.95 because of the MC error [Galindo-Garre et al., 2004]. This error tends to zero when number of simulations tend to infinity. As we implemented 2000 simulations, the MC error was equal to $\left(\frac{0.95\times0.05}{2000}\right)^{\frac{1}{2}} = 0.00487$ which means that coverage probabilities between 0.9451 and 0.9549 are in agreement with the nominal level of 95%. Therefore, the coverage probability 0.929, corresponding to β_2 in likelihood method, is clearly beyond the range above.

Though mean based bias and relative bias for the random effect estimate seems less in the likelihood method than in the corresponding Bayesian estimates, it is unacceptable as the average width of the 95% confidence interval (CI) is infinity due to upper CI limit being infinity at some simulations. However, the median width of 95% confidence intervals seems reasonable though higher than some of the Bayesian counterparts. Further, the likelihood-based method fails quite often when we want to estimate a 95% CI for the variance component σ_b^2 as this is not obtainable due to a non-positive definite covariance matrix. The intervals are not obtainable due to lack of estimated SEs which are equal to zero frequently in simulations with small sample (say n=6) particularly in likelihood methods. Thus, we excluded those problematic simulations during estimation of parameters under the likelihood method. This trouble also leads us to choose the Bayesian method assuming non-informative and slightly informative priors for the variance components.

It is relevant to discuss how we have chosen priors in this study. During selection of non-informative priors we followed Lambert et al. [2005] who used 13 different non-informative priors in a simulation study that demonstrates the potential influence of using prior distributions believed to be vague. A few of their non-informative or

weakly informative priors resemble ours. For fixed effects, we used non-informative normal priors as before in our real single study. However, all of their priors are not suitable for studying variance components in our study, for example, Pareto, half normal, uniform or normal priors as they produced highly biased estimates and had also low coverage probabilities and, therefore, these were excluded in our study.

Congdon [2007] suggests using normal priors with mean zero and large variance i.e. non-informative priors in the absence of prior knowledge. However, this might not be ideal when the sample size is small [Galindo-Garre et al., 2004]. In our case, simulation studies show that non-informative normal priors did not affect our coefficients relevant to fixed effects.

The use of non-informative priors for random effects are affected by having biased estimates. This is not unusual as clarified by Litière et al. [2008] that the estimate of the variability of random effects are always biased though the bias induced in the fixed effect parameters are small as long as the variability underlying the random effects distribution is small.

The results summarized in Table 2.20 are obtained using different non-informative and weakly informative priors. The number of Bayesian iterations was 5000 in each simulation where thinning was 10, burn-in period 2000, and number of chains was 4. Bias and MSE were computed with the formula as defined earlier in Section 2.10.1. For the width of Bayesian credible intervals, we simply average the width of 95% credible intervals obtained from S simulations. Median width is the median of respective widths of 95% CI both in likelihood and Bayesian methods. Median square error (MdSE) was calculated as the median of $(\hat{\beta}-\beta)^2$ [Galindo-Garre et al., 2004]. For confidence/credible intervals, we report coverage probabilities which represent the proportion of times that the true parameter lies within the 95% confidence intervals in likelihood method whereas in Bayesian case it is the proportion of times that the true parameter lies within 95% Bayesian credible intervals.

Table 2.20 shows that the estimates of fixed effects are fairly close to the true parameter values both in likelihood and Bayesian methods as bias or relative bias

Table 2.21: Simulated performance of likelihood and Bayesian estimates in fuel economy experiments. True parameter values: $\alpha = 32, \beta_2 = 1.4$, $\sigma_b^2 = 0.05$; Priors: $\alpha \sim N(0, 0.001)$, $\beta_2 \sim N(0, 0.001)$; simulation size=2000, n=40

Parameter	Mean/Posterior Mean			Median/Posterior Median			Coverage	Average width	Median width
	Bias	% Relative Bias	RMSE	Bias	% Relative Bias	RMdSE	Probability	of 95% CI	of 95% CI
MLE									
α	0.001	0.00	0.087	0.002	0.01	0.057	0.947	0.359	0.356
β_2	-0.002	-0.12	0.125	-0.001	-0.07	0.086	0.952	0.511	0.508
σ_b^2	0.001	2.17	0.026	-0.002	-3.50	0.018	0.976	∞	0.130
$\rho \sim$ Beta(1, 1)									
α	0.005	0.01	0.071	0.005	0.01	0.071	0.950	0.283	0.280
β_2	-0.004	-0.31	0.140	-0.004	-0.32	0.140	0.942	0.566	0.557
σ_b^2	0.002	3.86	0.026	-0.003	-6.93	0.023	0.956	0.111	0.105
$\rho \sim$ Beta(1.5, 1.5)									
α	0.007	0.02	0.070	0.007	0.02	0.070	0.959	0.287	0.290
β_2	-0.004	-0.26	0.145	-0.004	-0.30	0.145	0.939	0.574	0.571
σ_b^2	0.004	8.58	0.023	-0.001	-1.69	0.021	0.966	0.107	0.104
$\rho \sim$ Beta(2.5, 2.5)									
α	0.008	0.03	0.071	0.008	0.03	0.071	0.957	0.285	0.280
β_2	-0.004	-0.29	0.137	-0.005	-0.34	0.137	0.945	0.569	0.563
σ_b^2	0.004	7.54	0.020	-0.001	-1.84	0.019	0.980	0.098	0.095
$\rho \sim$ Beta(3, 3)									
α	0.008	0.02	0.073	0.007	0.02	0.073	0.948	0.287	0.290
β_2	-0.003	-0.25	0.139	-0.004	-0.28	0.139	0.963	0.573	0.570
σ_b^2	0.005	9.21	0.020	-0.000	-0.08	0.018	0.985	0.097	0.094
$1/\sigma_b^2 \sim$ Gamma(0.01, 0.001)									
α	0.004	0.01	0.071	0.005	0.02	0.071	0.983	0.323	0.320
β_2	-0.002	-0.17	0.142	-0.005	-0.36	0.142	0.969	0.615	0.610
σ_b^2	-0.035	-70.97	0.036	-0.043	-85.43	0.043	0.669	0.066	0.059
$1/\sigma_b^2 \sim$ Gamma(0.1, 0.1)									
α	0.014	0.04	0.073	0.015	0.05	0.073	0.992	0.376	0.370
β_2	0.001	0.09	0.141	0.002	0.14	0.141	0.989	0.733	0.729
σ_b^2	0.016	32.19	0.021	0.008	16.22	0.014	1.000	0.138	0.133

are very little. It is noted that the mean-based estimate of the day to day variation (σ_b^2) is relatively biased whereas median based estimates have less bias or relative bias in likelihood and Bayesian methods. When the priors $\rho \sim$ Beta$(1, 1)$, $\rho \sim$ Beta$(1.5, 1.5)$, $\rho \sim$ Beta$(2.5, 2.5)$, and $\rho \sim$ Beta$(3, 3)$ are considered the Bayesian median widths for all parameters are always less than average width of likelihood estimates. Particularly, the width of the likelihood based 95% CI is infinity because of having the upper bound infinity in some trials of the simulation process. $\rho \sim$ Beta(1,1) which is equivalent to $\rho \sim$ U(0,1) should not be used to estimate variance component σ_b^2 as simulation results shows that the relative bias is approximately 243.87% , while bias is less in the case of a weakly informative prior $\rho \sim$ Beta(2.5, 2.5) or $\rho \sim$ Beta(3, 3). However, relative bias corresponding to σ_b^2 is minimal if we consider median based estimates. The gamma priors are not suitable for estimating fixed effects as well as variance components as they produce severely biased estimates for the variance component and high coverage probabilities for all parameters, particularly when sample size is small, for instance, n=6. Coverage probabilities are acceptable for fixed effects as these are close to nominal coverage, however for the variance component σ_b^2, they appeared as over coverage which might lead to erroneous conclusions about σ_b^2.

In Table 2.21, the sample size has been increased to 40. Now results are improved in the indices of bias, root mean squared error (RMSE) or root median squared error (RMdSE), coverage probability, average and median widths. It seems that both MLE and Bayesian methods, particularly with priors $\rho \sim$Beta(1, 1) or $\rho \sim$ Beta(1.5, 1.5), produce good results in terms of relative bias. However, the Bayesian estimates, particularly with priors $\rho \sim$Beta(1, 1) to $\rho \sim$Beta(3, 3), performed better than MLE in terms of average and median width of 95% CIs. All of these sets provide precise fixed effect estimates. Though the variance component (σ_b^2) is estimated well in the ML method its average width of 95% CI is extreme, whereas median width in Bayesian method is reasonable. Further, in the likelihood method there were 17.58% cases in the iterations where SE estimate of σ_b^2 was 0 and many cases it was infinity, therefore, average width of 95% CI was infinity. The cases where SE estimate of σ_b^2 were not available has been excluded from the sample during estimation.

Table 2.22: Performance of likelihood and Bayesian estimates in fuel economy experiments considering different size of simulations. True parameter values: $\alpha = 32, \beta_2 = 1.4$, $\sigma_b^2 = 0.05$; Priors: $\alpha \sim N(0, 0.001)$, $\beta_2 \sim N(0, 0.001)$; n=40

Parameter	Mean/Posterior Mean			Median/Posterior Median			Coverage	Average width	Median width
	Bias	% Relative Bias	RMSE	Bias	% Relative Bias	RMdSE	Probability	of 95% CI	of 95% CI
MLE, S=2000									
α	-0.001	-0.00	0.085	0.000	0.00	0.059	0.955	0.357	0.354
β_2	-0.000	-0.03	0.122	0.000	0.03	0.082	0.950	0.508	0.504
σ_b^2	0.001	1.20	0.025	-0.002	-4.99	0.017	0.976	∞	0.127
MLE, S=5000									
α	-0.001	-0.00	0.088	-0.000	-0.00	0.060	0.948	0.358	0.356
β_2	0.002	0.13	0.124	0.004	0.29	0.083	0.949	0.510	0.507
σ_b^2	0.001	1.60	0.026	-0.002	-3.25	0.018	0.974	∞	0.130
$\rho \sim$ Beta(2.5, 2.5), S=2000									
α	0.008	0.03	0.071	0.008	0.03	0.071	0.957	0.285	0.280
β_2	-0.004	-0.29	0.137	-0.005	-0.34	0.137	0.945	0.569	0.563
σ_b^2	0.004	7.54	0.020	-0.001	-1.84	0.019	0.980	0.098	0.095
$\rho \sim$ Beta(2.5, 2.5), S=5000									
α	0.001	0.00	0.071	0.001	0.00	0.071	0.953	0.284	0.280
β_2	-0.003	-0.23	0.142	-0.003	-0.20	0.142	0.945	0.571	0.568
σ_b^2	0.004	7.72	0.020	-0.001	-1.62	0.019	0.978	0.098	0.096

Table 2.23: Robustness of likelihood and Bayesian estimates in fuel economy experiments with different values of σ^2 and σ_b^2 where true parameter values are $\alpha = 32$ and $\beta_2 = 1.4$, and priors given $\alpha \sim N(0, 0.001)$ and $\beta_2 \sim N(0, 0.001)$; simulation size=2000, n=40

Parameter	Mean/Posterior Mean			Median/Posterior Median			Coverage	Average width	Median width
	Bias	% Relative Bias	RMSE	Bias	% Relative Bias	RMdSE	Probability	of 95% CI	of 95% CI
MLE, σ^2=0.05 and σ_b^2=0.05									
α	0.001	0.00	0.087	0.002	0.01	0.057	0.947	0.359	0.356
β_2	-0.002	-0.12	0.125	-0.001	-0.07	0.086	0.952	0.511	0.508
σ_b^2	0.001	2.17	0.026	-0.002	-3.50	0.018	0.976	∞	0.130
MLE, σ^2=0.05 and σ_b^2=1									
α	-0.010	-0.03	0.324	-0.016	-0.05	0.213	0.949	1.320	1.317
β_2	0.006	0.44	0.463	0.012	0.89	0.319	0.946	1.880	1.876
σ_b^2	0.005	0.52	0.349	-0.029	-2.90	0.231	0.934	1.450	1.400
MLE, σ^2=1 and σ_b^2=0.05									
α	-0.005	-0.01	0.231	-0.009	-0.03	0.156	0.965	1.019	1.010
β_2	0.008	0.55	0.332	0.010	0.73	0.231	0.966	1.451	1.438
σ_b^2	0.113	225.25	0.208	0.067	133.91	0.067	0.784	∞	6.195
MLE, σ^2=1 and σ_b^2=1									
α	-0.006	-0.02	0.396	-0.015	-0.05	0.267	0.938	1.597	1.590
β_2	0.007	0.52	0.565	0.018	1.31	0.369	0.944	2.275	2.264
σ_b^2	0.009	0.93	0.505	-0.028	-2.82	0.340	0.983	∞	2.572
$\rho \sim$ Beta(2.5, 2.5), σ^2=0.05 and σ_b^2=0.05									
α	0.008	0.03	0.071	0.008	0.03	0.071	0.957	0.285	0.280
β_2	-0.004	-0.29	0.137	-0.005	-0.34	0.137	0.945	0.569	0.563
σ_b^2	0.004	7.54	0.020	-0.001	-1.84	0.019	0.980	0.098	0.095
$\rho \sim$ Beta(2.5, 2.5), σ^2=0.05 and σ_b^2=1									
α	0.029	0.09	0.260	0.027	0.09	0.260	0.917	0.936	0.930
β_2	-0.034	-2.41	0.519	-0.030	-2.16	0.518	0.917	1.884	1.861
σ_b^2	-0.138	-13.84	0.327	-0.192	-19.22	0.339	0.869	1.114	1.066
$\rho \sim$ Beta(2.5, 2.5), σ^2=1 and σ_b^2=0.05									
α	0.029	0.09	0.188	0.027	0.08	0.188	0.978	0.890	0.890
β_2	-0.027	-1.95	0.375	-0.026	-1.85	0.375	0.974	1.789	1.788
σ_b^2	0.314	627.05	0.335	0.267	534.04	0.288	0.264	0.843	0.823
$\rho \sim$ Beta(2.5, 2.5), σ^2=1 and σ_b^2=1									
α	0.053	0.16	0.325	0.049	0.15	0.325	0.950	1.264	1.260
β_2	-0.051	-3.64	0.643	-0.046	-3.31	0.643	0.943	2.548	2.538
σ_b^2	0.060	6.04	0.400	-0.033	-3.25	0.374	0.986	1.956	1.892

2.10.4 Robustness of Likelihood and Bayesian Estimators

Table 2.20 and Table 2.21 show that bias and relative bias are small and similar for relative to the estimates of fixed effects. However, bias and relative bias are reduced greatly for both likelihood and Bayesian methods when we increase the sample size from 6 to 40.

Performance of likelihood and Bayesian estimates are observed in all tables with respect to estimation methods or priors. It is essential to see how estimates vary with respect to various parameters, for instance, with respect to simulation sizes (S), sample sizes (n), with different settings of σ^2 and σ_b^2. The estimates are improved by having less bias or relative bias with the increasing sample sizes. For example, we see that the bias or relative bias is more in Table 2.20 with sample size 6 and less in Table 2.21 with sample size 40. Table 2.22 shows that fixed and random effects are not varying much with respect to the number of simulations. If we run 5000 simulations instead of 2000, there are no substantial changes in estimated bias, relative bias, RMSE, coverage probabilities and 95% widths of CIs. So, we choose simulation size as 2000 for all successive trials in this chapter.

Table 2.23 describes the robustness of likelihood and Bayesian estimates under different choices of σ^2 and σ_b^2. The likelihood and Bayesian estimates perform well except the choice of $\sigma^2 = 1$ and $\sigma_b^2 = 0.05$. Again, the performance of mean or median based likelihood estimates are good except for σ_b^2 as corresponding average width of 95% CI could be infinity which indicates that upper limit of 95% CI is infinity in some simulations. However, corresponding Bayesian estimate of σ_b^2 is more robust as none of the choices of σ^2 and σ_b^2 results average width of 95% CI being infinity.

2.10.5 Simulation Studies on Round Robin Experiments

There were 16 laboratories in the real round robin experiments and the Bayesian estimates corresponding to fuel A and fuel B were acceptable as supported by convergence diagnostics. Though the pattern of estimates corresponding to fuel A and fuel B were different (see Table 2.11 and Table 2.12), they were usual in their own scale. In the first simulation study, we take the number of laboratories as

Table 2.24: Simulated performance of likelihood and Bayesian estimates in round robin experiments. True parameter values: α=22, σ^2=16 σ_L^2=5, r=11.2, R=12.831 ; simulation size=2000, n=20

Parameter	Mean/Posterior Mean			Median/Posterior Median			Coverage	Average width	Median width
	Bias	% Relative Bias	RMSE	Bias	% Relative Bias	RMdSE	Probability	of 95% CI	of 95% CI
MLE									
α	-0.013	-0.06	1.156	-0.010	-0.05	0.778	0.960	5.182	5.103
σ^2	-1.257	-7.86	6.331	-1.989	-12.43	4.486	0.918	28.736	26.414
σ_L^2	1.712	34.23	6.151	0.512	10.24	3.894	0.944	∞	71.215
r	-0.690	-6.16	2.369	-0.719	-6.42	1.620	0.918	9.195	9.051
R	-0.042	-0.33	2.153	-0.118	-0.92	1.497	1.000	∞	21.448
$\rho \sim$ Beta(1, 1)									
α	0.047	0.21	1.154	0.043	0.20	1.155	0.963	5.198	5.121
σ^2	0.580	3.62	6.075	-0.867	-5.42	5.672	0.949	26.073	24.951
σ_L^2	4.389	87.78	6.708	2.116	42.32	4.833	0.987	30.740	28.396
r	-0.204	-1.82	2.059	-0.496	-4.42	2.074	0.949	8.412	8.352
R	0.975	7.60	2.537	0.550	4.29	2.330	0.946	10.407	10.288
$\rho \sim$ Beta(1.5, 1.5)									
α	-0.022	-0.10	1.129	-0.024	-0.11	1.130	0.975	5.320	5.247
σ^2	-0.117	-0.73	5.636	-1.491	-9.32	5.422	0.946	24.748	23.870
σ_L^2	5.487	109.74	7.274	3.292	65.83	5.271	0.979	31.155	29.198
r	-0.421	-3.76	1.986	-0.706	-6.30	2.031	0.946	8.161	8.103
R	1.099	8.56	2.516	0.677	5.28	2.298	0.949	10.480	10.386
$\rho \sim$ Beta(2.5, 2.5)									
α	0.006	0.03	1.170	0.005	0.02	1.170	0.966	5.333	5.302
σ^2	-0.827	-5.17	5.300	-2.104	-13.15	5.284	0.935	22.987	22.252
σ_L^2	6.128	122.55	7.543	4.108	82.17	5.604	0.959	29.894	28.887
r	-0.646	-5.77	1.966	-0.922	-8.23	2.041	0.935	7.754	7.727
R	1.088	8.48	2.484	0.676	5.27	2.270	0.954	10.443	10.406
$\rho \sim$ Beta(3, 3)									
α	0.015	0.07	1.163	0.015	0.07	1.163	0.967	5.375	5.316
σ^2	-0.896	-5.60	5.296	-2.162	-13.51	5.279	0.940	22.778	21.827
σ_L^2	6.516	130.31	7.824	4.545	90.90	5.897	0.947	29.625	28.198
r	-0.666	-5.95	1.960	-0.941	-8.40	2.033	0.940	7.694	7.627
R	1.171	9.13	2.555	0.761	5.93	2.333	0.950	10.462	10.382
$1/\sigma_L^2 \sim$ Gamma(0.01, 0.001)									
α	-3.95e+05	-1.79e+06	1.13e+06	-0.179	-0.81	1.129	0.956	1.70e+06	5.118
σ^2	3.188	19.92	7.801	1.489	9.31	6.891	0.955	31.580	30.158
σ_L^2	6.39e+12	1.27e+14	2.10e+13	-1.516	-30.31	6.267	0.963	4.24e+13	25.309
r	5.87e-01	5.25	2.357	0.284	2.53	2.306	0.956	9.527	9.366
R	1.19e+06	9.35e+06	3.43e+06	0.478	3.72	2.185	0.968	6.38e+06	10.742
$1/\sigma_L^2 \sim$ Gamma(0.1, 0.1)									
α	-0.032	-0.15	1.152	-0.037	-0.17	1.142	0.949	4.873	4.757
σ^2	2.557	15.98	7.246	0.859	5.37	6.408	0.956	30.702	29.459
σ_L^2	1.434	28.68	10.999	-1.242	-24.83	4.659	0.990	28.808	22.300
r	0.397	3.54	2.247	0.084	0.75	2.202	0.956	9.396	9.310
R	0.676	5.27	2.335	0.273	2.13	2.150	0.958	10.084	9.840

Table 2.25: Simulated performance of likelihood and Bayesian estimates in round robin experiments. True parameter values: α=22, σ^2=16 σ_L^2=5, r=11.2, R=12.831 ; simulation size=2000, n=40

Parameter	Mean/Posterior Mean			Median/Posterior Median			Coverage Probability	Average width of 95% CI	Median width of 95% CI
	Bias	% Relative Bias	RMSE	Bias	% Relative Bias	RMdSE			
MLE									
α	-0.032	-0.14	0.795	-0.027	-0.12	0.533	0.958	3.411	3.378
σ^2	-0.744	-4.65	4.706	-1.361	-8.51	3.341	0.930	19.870	18.836
σ_L^2	1.032	20.64	4.527	0.358	7.15	3.001	0.928	∞	33.418
r	-0.389	-3.48	1.699	-0.487	-4.35	1.196	0.930	6.715	6.616
R	0.003	0.02	1.477	-0.010	-0.08	1.007	0.999	∞	13.332
$\rho \sim$ Beta(1, 1)									
α	0.011	0.05	0.811	0.012	0.06	0.811	0.953	3.367	3.340
σ^2	0.148	0.92	4.186	-0.563	-3.52	4.102	0.957	17.869	17.600
σ_L^2	1.917	38.34	4.032	0.939	18.78	3.509	0.987	18.085	17.285
r	-0.154	-1.37	1.464	-0.295	-2.63	1.482	0.957	6.044	6.034
R	0.422	3.29	1.595	0.243	1.89	1.533	0.946	6.482	6.432
$\rho \sim$ Beta(1.5, 1.5)									
α	-0.002	-0.01	0.790	-0.002	-0.01	0.792	0.966	3.429	3.411
σ^2	-0.401	-2.50	4.029	-1.074	-6.71	4.034	0.956	16.972	16.634
σ_L^2	2.768	55.36	4.375	1.820	36.40	3.683	0.977	18.348	17.739
r	-0.336	-3.00	1.451	-0.474	-4.23	1.484	0.957	5.839	5.829
R	0.510	3.97	1.615	0.330	2.57	1.545	0.951	6.550	6.512
$\rho \sim$ Beta(2.5, 2.5)									
α	0.018	0.08	0.809	0.018	0.08	0.810	0.961	3.497	3.472
σ^2	-1.031	-6.44	3.942	-1.661	-10.38	4.040	0.938	15.885	15.502
σ_L^2	3.682	73.63	4.805	2.774	55.48	4.020	0.961	18.093	17.626
r	-0.551	-4.92	1.473	-0.685	-6.11	1.521	0.938	5.579	5.553
R	0.595	4.64	1.611	0.417	3.25	1.535	0.958	6.594	6.571
$\rho \sim$ Beta(3, 3)									
α	-0.031	-0.14	0.812	-0.030	-0.14	0.812	0.960	3.523	3.520
σ^2	-1.237	-7.73	3.976	-1.852	-11.58	4.095	0.925	15.537	15.315
σ_L^2	4.050	81.00	5.057	3.160	63.21	4.254	0.943	17.995	17.584
r	-0.625	-5.58	1.512	-0.757	-6.76	1.563	0.925	5.492	5.498
R	0.637	4.97	1.665	0.460	3.58	1.586	0.945	6.614	6.609
$1/\sigma_L^2 \sim$ Gamma(0.01, 0.001)									
α	0.003	0.01	0.789	0.006	0.03	0.789	0.935	3.053	3.020
σ^2	3.578	22.36	6.294	2.830	17.69	5.855	0.924	21.836	21.181
σ_L^2	-2.216	-44.32	4.097	-3.550	-71.00	4.875	0.816	12.594	10.356
r	0.949	8.48	1.904	0.834	7.45	1.870	0.924	6.794	6.682
R	0.221	1.72	1.569	0.054	0.42	1.531	0.942	6.143	6.067
$1/\sigma_L^2 \sim$ Gamma(0.1, 0.1)									
α	0.011	0.05	0.805	0.014	0.06	0.806	0.944	3.180	3.140
σ^2	2.254	14.09	5.301	1.440	9.00	4.932	0.952	21.037	20.596
σ_L^2	-0.487	-9.75	3.703	-1.715	-34.30	3.986	0.963	15.473	13.656
r	0.528	4.72	1.664	0.384	3.43	1.637	0.953	6.727	6.650
R	0.340	2.65	1.562	0.169	1.32	1.511	0.950	6.290	6.203

20, true parameters as α=22, σ^2=16 and σ_L^2=5 that determine r=11.2 and R=12.831 using formula (2.14) and the number of simulations as 2000. The prior for α was non-informative normal and for ρ a range of beta priors was used and also we tested non-informative, weakly informative gamma priors for the precision parameter $1/\sigma_L^2$, corresponding to lab-to-lab variability.

We see in Table 2.24 that the mean-based bias, relative bias and RMSE or RMdSE are more or less similar both in likelihood and Bayesian methods. Though bias, relative bias and RMSE seem better in median based likelihood estimates, the quality of interval estimates are not better than those of their Bayesian counterparts. Particularly repeatability (r) has under coverage 0.918 and reproducibility (R) has over-coverage at 1 which means that 100% of the 95% confidence intervals include the true value. Also, the average width of σ_L^2 is infinity and hence the average width of R is infinity. Average and median width of 95% CIs are higher than their Bayesian analogues. Among the Bayesian estimates the set with prior $\rho \sim$Beta(1, 1) or $1/\sigma_L^2 \sim$ Gamma(0.1, 0.1) could be a better option because these produce minimum bias, RMSE or RMdSE, small width of 95% CIs along with reasonable coverage probabilities.

We have increased the number of laboratories from 20 to 40 to see the improvement of likelihood and Bayesian estimates which are presented in Table 2.25. The true parameter values were as in Table 2.24. We notice that parameter estimates both in likelihood and Bayesian methods have been improved in all indices. Bias, relative bias and RMSE are less in Table 2.25 with the increase of sample size in comparison to Table 2.24. In the likelihood method still the average width is infinity due to having some upper limits of 95% CIs of σ_L^2 as infinity. Among Bayesian estimators the sets with $\rho \sim$ Beta(1, 1) and with $1/\sigma_L^2 \sim$ Gamma(0.1, 0.1) performed better than others by providing less bias along with less RMSE, minimum width of credible intervals and with sensible coverage probabilities as particularly the likelihood method yield coverage probabilities lower than the 95% nominal level in round robin experiments.

2.10.6 Kernel Density of Simulated Estimates

Kernel density can be used to visualize the sampling distribution of simulated estimates. It is a non-parametric method to estimate the probability density function of a random variable. Kernel density estimates are similar to histograms, but can be endowed with properties such as smoothness or continuity by using a suitable kernel.

Figure 2.10 shows the kernel densities of the estimates from the simulated fuel economy experiments. During generation of samples true parameter values were $\alpha = 32$, $\beta_2 = 1.4$, $\sigma_b^2 = 0.05$, sample size was 40 and the priors in the Bayesian analysis were $\alpha \sim N(0, 0.001)$, $\beta_2 \sim N(0, 0.001)$, and $\rho \sim$Beta(1, 1). In the figure, the shapes of kernel densities corresponding to base fuel (α) and fuel difference (β_2) are approximately normal both in likelihood and Bayesian methods. The kernel densities indicate that the distributions of day to day variability (σ_b^2) under likelihood and Bayesian methods are positively skewed. There could have been more zero estimates for day to day variability in the case of the likelihood method than the Bayesian method. Therefore, it is understandable that there could have been many zero estimates of day to day variation in real fuel economy experiments.

To have a better sense of why the density plots have higher bumps at certain places, we look at rug plots (see below the kernel density plots in Figure 2.11). A rug plot is a plot of tick marks along the horizontal axis indicating where the data are located. Clearly, there are more data in the neighbourhood between 31.5 and 32.5 where highest 'bump' is located for base fuel (α) both in likelihood and Bayesian simulation studies. The kernel densities and rug plots for fuel difference (β_2) also have the same pattern in both methods. However, day to day variability (σ_b^2) has different distributional patterns for likelihood and Bayesian estimates.

It seems that likelihood and Bayesian methods perform approximately equally well. Yet, there were 1638 simulations out of 2000 that were successfully obtained likelihood estimates because of arising non-positive definite variance covariance matrices in the likelihood method, whereas the Bayesian method did not encounter such problems. This is the main advantage of using Bayesian methods in fuel economy experiments. Also, in terms of coverage probabilities and width of 95% confidence/credible intervals Bayesian methods are better than likelihood based

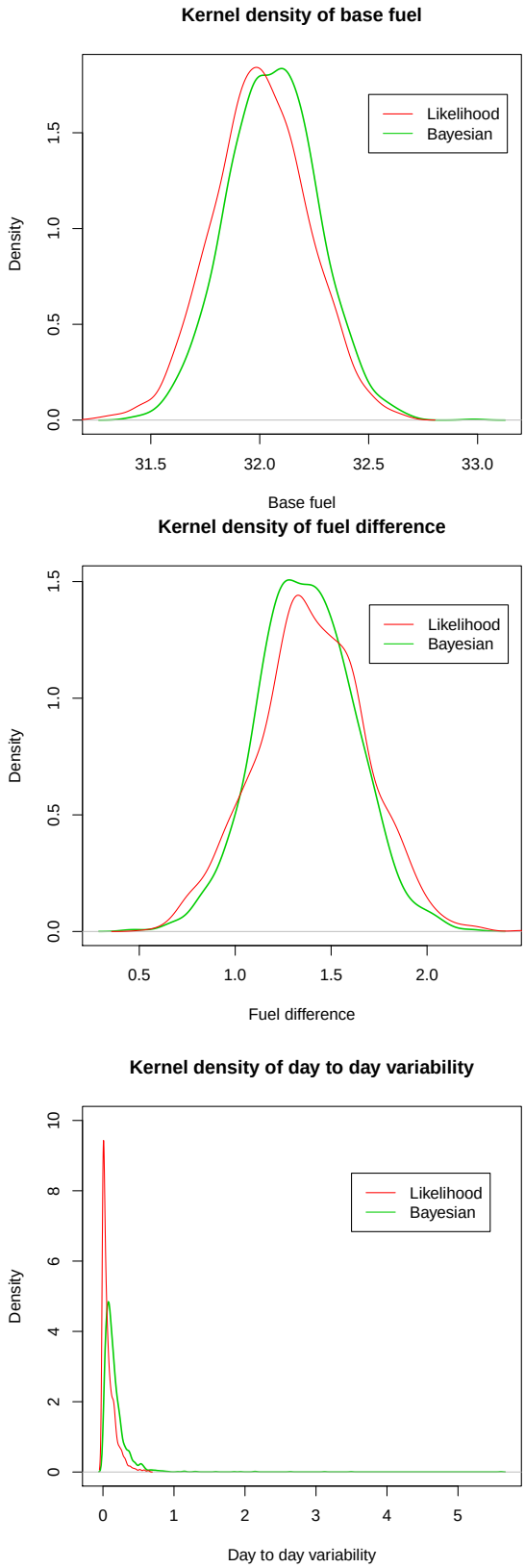

Figure 2.10: Kernel density of simulated parameter estimates in fuel economy experiments; base fuel (α)(top), fuel difference (β_2) (middle), day to day variability (σ_b^2) (bottom)

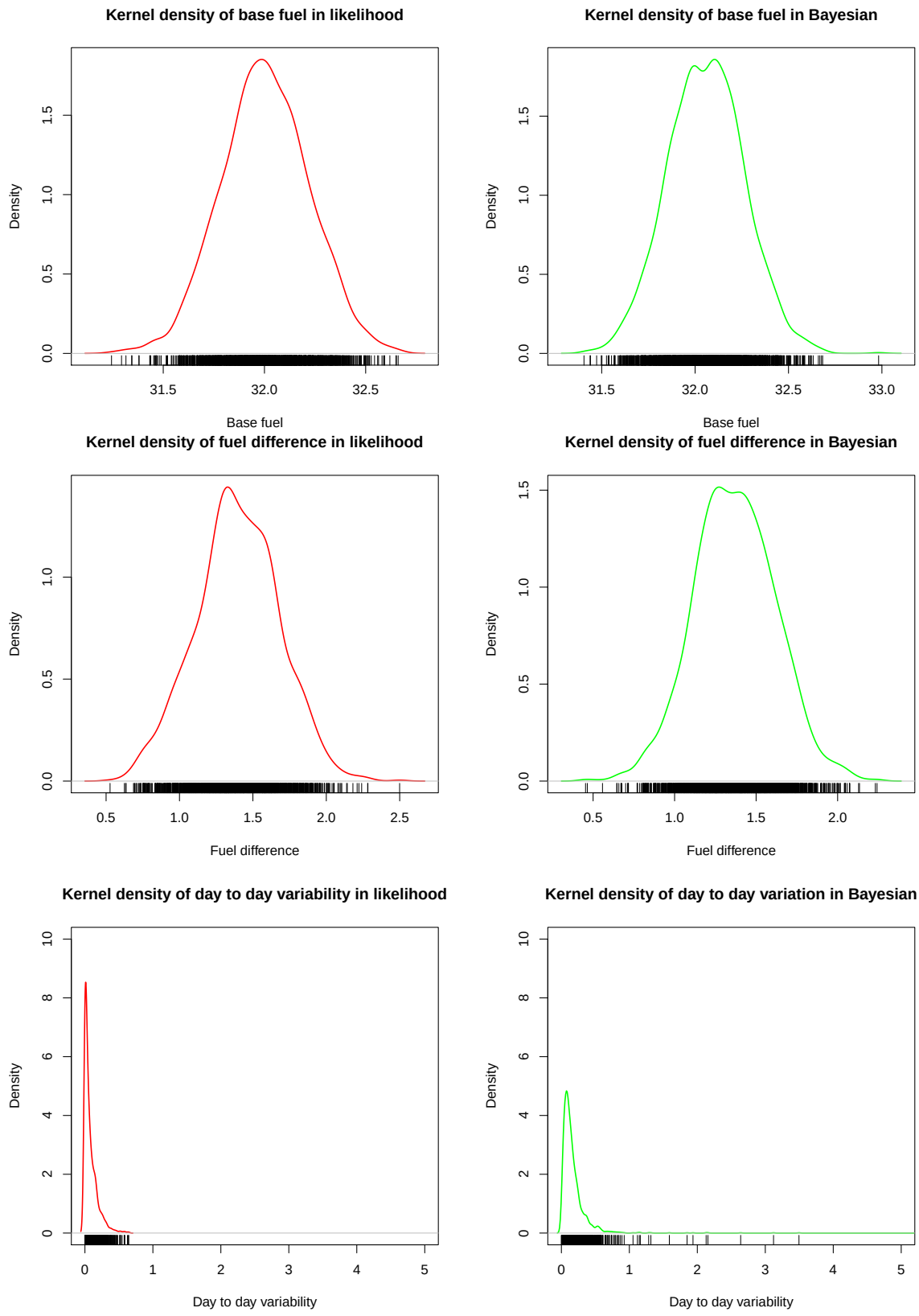

Figure 2.11: Kernel density of simulated parameter estimates in fuel economy experiments; Base fuel (α)(top), Fuel difference (β_2) (middle), Day to day variability (σ_b^2) (bottom)

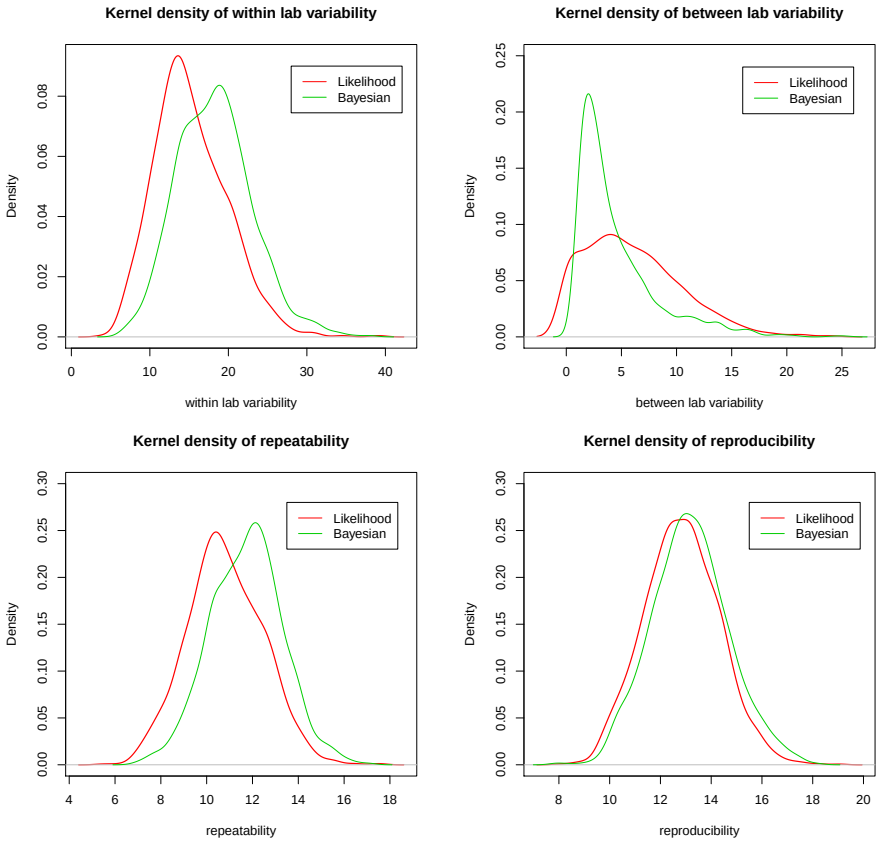

Figure 2.12: Kernel density of simulated parameter estimates in round robin experiments; within lab variability (top left), between lab variability (top right), repeatability (bottom left), reproducibility (bottom right)

methods as they provide good coverage estimates which are close to the 95% nominal confidence limit and have less width in comparison to likelihood method (see Table 2.21).

The kernel densities of four parameters within lab variability (σ^2), between lab variability (σ_L^2), repeatability (r) and reproducibility (R) in round robin experiments are shown in Figure 2.12. The likelihood and Bayesian kernel density curves are overlaid on each other. In the figure we see each set of kernel density curves are similar both in likelihood and Bayesian methods except density curves for between lab variability where the likelihood kernel density is wider than the Bayesian counterpart. Thus Bayesian estimates are more stable than the likelihood estimates. Though the kernel density of between lab variability was not usual, the reproducibility which is function of σ^2 and σ_L^2 was not abnormal. The kernel densities for R nearly coincide for likelihood and Bayesian methods. However, again there were 1797 simulations successful out of 2000 simulations as the rest of them failed in the likelihood parameter estimation due to arising non-positive definite matrices at some points of simulation in likelihood method. Figure 2.13 provides additional information about raw estimates through rug plots under the kernel densities. It is evident from this figure that the density plots of within lab variability (σ^2) and repeatability (r) are similar as repeatability is a function of within lab variability.

2.11 Conclusion

In this Chapter, we have applied Bayesian methods in the fuel efficiency experiment which is a novel work in this field. This study has enabled variance components estimation that was poorly estimable in classical methods due to small number of strata (groups). Likelihood-based REML and ML methods have estimated the variance component due to days approximately as zero which is unrealistic. Therefore, we have implemented Bayesian techniques assuming some priors to determine the day-to-day variance component. However, the Bayesian estimate of the variance component is inflated as evidenced by Lambert et al. [2005] who noted that MCMC methods provide the estimate of variance parameter as upwardly biased. Further, as the standard asymptotic theory breaks down in case of deriving confidence interval for variance component in likelihood method, we have compared the Bayesian 95% credible interval of variance component with the confidence intervals based on profile likelihood and bootstrap methods. To evaluate the quality

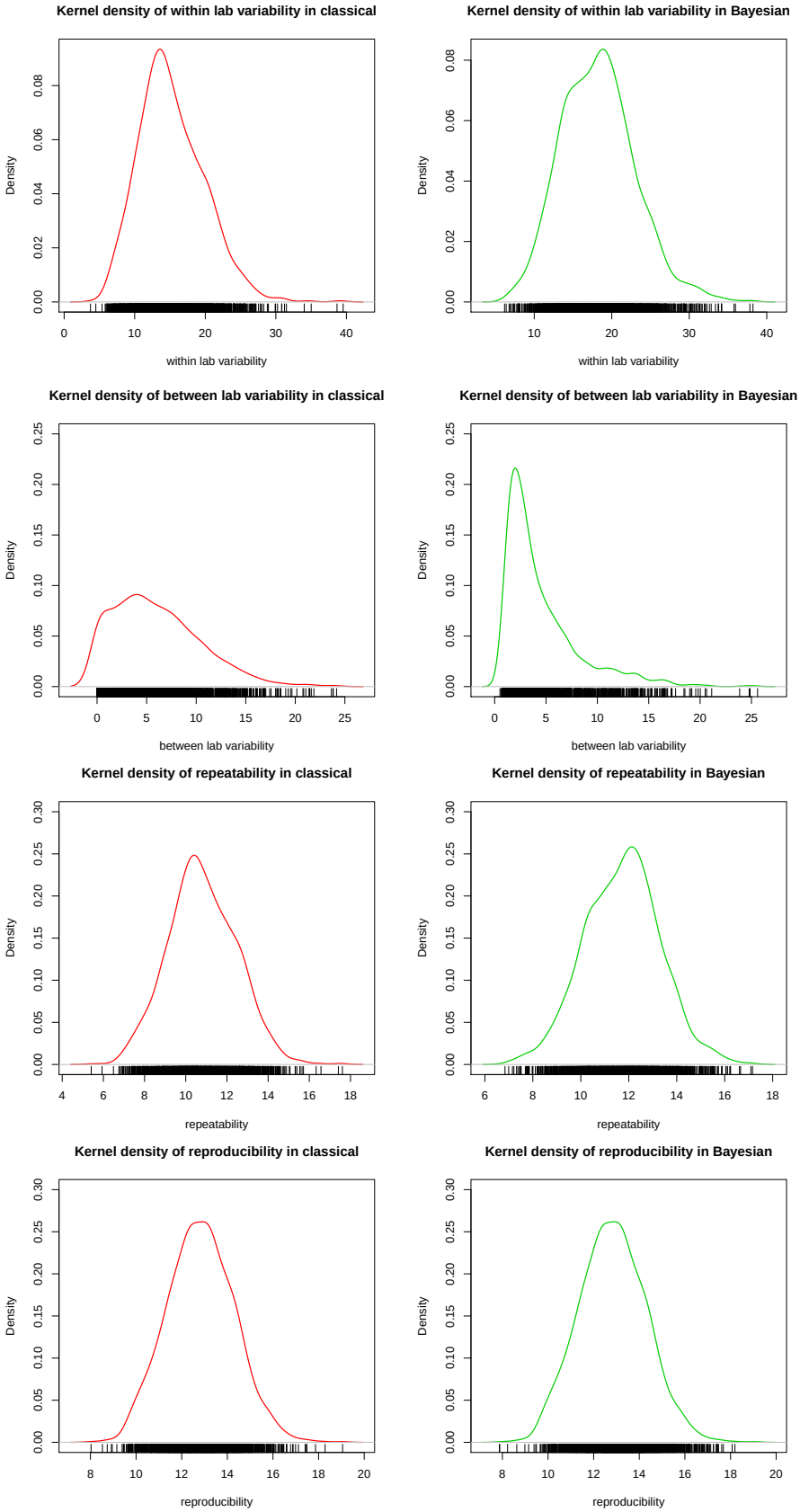

Figure 2.13: Kernel density of simulated parameter estimates in round robin experiments; within lab variability (top), between lab variability (second from top), repeatability (third from top), reproducibility (bottom)

of Bayesian as well as likelihood estimates, we have performed simulation studies. In this regard, the frequentist properties of bias, accuracy, and coverage of the parameter estimates have been investigated.

The analysis of real fuel economy data shows that the fixed effect estimates are similar both in classical and Bayesian methods. However, the estimates of variance components differ substantially. For instance, day-to-day variation in the model corresponding to contrast T-B was close to 0 in classical method, whereas the Bayesian estimate was 0.059 (see Table 2.5, Table 2.14). Though Bayesian methods ensure that the variance component is not estimated to be 0, but the Bayesian estimation is not free from criticism as it suffers from overestimation of the point estimates. Therefore, the Bayesian estimate of the variance component has been compared with profile likelihood and bootstrap based intervals which ensure that the Bayesian estimate is not absurd because the Bayesian point estimate of day to day variance was 0.059 which lies within the profile likelihood based interval (0.000, 0.099) shown in Table 2.18.

In the analysis of nested model (2.12), 'days' were nested under 'weeks'. The experiment was conducted in two weeks containing an interval of 4-5 days between them. As the Bayesian analysis has revealed that there is no effect of week on the responses, it is not essential to keep a break in the middle of experiment. Therefore, experimenters can save time by not keeping the provision of a 4-5 days gap during the fuel economy experiments.

To verify the correct estimation of the posterior distributions we have used convergence diagnostics namely MC error, kernel density estimation, autocorrelation, trace-plot, history plot and Gelman-Rubin statistic. All the Bayesian models in fuel economy and round robin experiments have passed the convergence diagnostics.

A simulation study was performed to evaluate which method - likelihood or Bayesian produces the best estimates. The performance of the simulated point estimates were assessed by bias or mean and median squared error and interval estimates were assessed by the coverage probabilities and mean and median width of 95% confidence or credible intervals. A good point estimator has small bias and small mean/median squared error and a good confidence interval has small width under

the condition that its coverage probability is at least 0.95 [Galindo-Garre et al., 2004]. Keeping these criteria in mind, the simulation results on fuel economy and round robin experiments can be summarised as follows.

In small fuel economy experiment, say n=6 in Table 2.20, it seems that fixed effect estimates are reasonable in terms of bias, coverage probability and width both in likelihood and Bayesian methods. However, the mean width of 95% is infinity as at least one width in the simulated samples is infinity due to the upper limit of that interval being infinity. Therefore, a mean based estimate of the variance component (σ_b^2) is not acceptable. With respect to median based estimates a Bayesian approach performed better, particularly the set of priors with ρ ∼Beta(1, 1) and ρ ∼Beta(1.5, 1.5) performed better than the estimates of the likelihood method. For fixed effects the main concern is to estimate fuel difference (β_2). Though likelihood and Bayesian estimates of (β_2) are fairly close, the corresponding coverage probability in the likelihood method is slightly below the nominal level 0.95. However, for small samples, it seems that likelihood method underestimates the variance component (σ_b^2) and Bayesian method overestimates it. From our intuition we conclude that none of the estimates of (σ_b^2) obtained in likelihood or Bayesian method is accurate, rather perhaps it is in between likelihood and Bayesian estimates. The mean or median width of the Bayesian credible intervals are smaller than the median width of the classical confidence intervals particularly when the set of priors with ρ ∼Beta(1, 1) to ρ ∼Beta(3, 3). Among the priors the set with ρ ∼Beta(3, 3) performs the best for small sample in fuel economy experiment. When the sample size is increased from 6 to 40, the estimates get improved by providing less bias, close to desired coverage probabilities, and smaller widths of intervals (see Table 2.21). However, the likelihood and Bayesian methods perform fairly close to each other except the pitfall in average width of 95% confidence interval of σ_b^2 in likelihood method.

In round robin experiments, the estimates of fixed effects are similar for likelihood and Bayesian methods, but the likelihood estimate of lab to lab variability (σ_L^2) is zero, whereas the Bayesian estimate of lab to lab variability is 11.517. Though we understand that the Bayesian estimate is upwardly biased, but impressively the Bayesian estimate is still within the profile likelihood and bootstrap based confidence intervals (see Table 2.19). Yet, the quality of the likelihood and

Bayesian estimates under round robin experiments have been assessed by a simulation study. The simulated performance of likelihood and Bayesian estimates are shown in Table 2.24 and Table 2.25. The comparison indices are bias or relative bias, coverage probabilities, mean/median width of intervals. The coverage probabilities are unusual in likelihood method i.e. either below or over coverage than the nominal level 0.95. In the Bayesian case, with different sets of priors the estimates seems reasonable with acceptable coverage probabilities. However, a non-informative gamma prior e.g. $1/\sigma_L^2 \sim$ Gamma(0.01, 0.001) is not suitable for studying round robin experiments. We notice that there is little differences in results while priors $\rho \sim$Beta(1, 1), $\rho \sim$Beta(1.5, 1.5), and $\rho \sim$Beta(2.5, 2.5) are taken into account. Perhaps we can recommend the prior sets with $\rho \sim$Beta(1.5, 1.5) or $\rho \sim$Beta(2.5, 2.5) for Bayesian analysis of round robin experiments. The Bayesian analysis would be more appropriate as the likelihood method has the shortcoming of having infinite average width of confidence intervals.

Throughout the simulation studies a package R2WinBUGS has been used to call WinBUGS from R. Basically R2WinBUGS makes use of batch mode feature and provides tools to call WinBUGS directly after data manipulation in R. After the WinBUGS process had finished, it was possible to work with the results importing them back into R again, for example, essential posterior summaries were saved in R for further processing and to prepare outputs for simulation studies.

At the end, we might conclude that the newly applied Bayesian methods in fuel efficiency field appeared to be a strong competitor with usual classical basis of analysis for Shell statistical research unit. In the case of small experiments where there is a chance of computational failure in classical methods, the Bayesian methods can be used with the appropriate choice of priors. These techniques of analyzing fuel efficiency can be replicated to other industrial experiments with small number of groups.

Chapter 3

Bayesian Analysis of Categorical Data from Multi-Stratum Experiments

3.1 Introduction

Residual maximum likelihood (REML) method is used to estimate random effects and empirical generalized least squares (GLS) to estimate fixed effects in linear mixed models. However, REML-GLS estimation can give misleading conclusions if there are few main plots in a non-orthogonal split-plot design. Some of the variance components were estimated to be 0, perhaps due to inadequate number of whole plots during the analysis of data from a polypropylene experiment. To estimate the variance components properly Goos and Gilmour [2012] suggested the possibility of doing a Bayesian analysis assuming informative priors for the variance components, though they did not implement this suggestion.

In this chapter, we have implemented Bayesian methods considering some informative priors for variance components and noninformative priors for the fixed effects in multi-stratum and split-plot design settings with binary and ordered categorical responses. The added value of the Bayesian approach is that it enables variance components estimation when these are inestimable or poorly estimated in classical methods, thus leading to more reasonable standard errors and inferences for the fixed effects of the treatment factors. We illustrate our approach using the data

from the polypropylene experiment described in Section 3.4 onwards.

3.2 Models to be Used in the Analysis

Mixed models are used to analyze multi-stratum and split-plot designs as each stratum may have random effects on the responses [Letsinger et al., 1996]. We have two types of models namely mixed binary logit and mixed cumulative logit models for analyzing categorical data from the polypropylene experiment, depending on how the responses are categorized.

If responses are binary then binary logit model is appropriate for analysis. When some factors have random effects on the responses we use mixed binary logit model. For example, the mixed binary logistic mixed model can be written as

$$\begin{aligned} Y_{ij} \mid \delta_i, \epsilon_{ij} &\sim \text{Bernoulli}(P_{ij}) \\ \text{logit}(P_{ij}) &= \beta_0 + \mathbf{x}'_{ij}\boldsymbol{\beta} + \delta_i + \epsilon_{ij} \end{aligned} \tag{3.1}$$

where Y_{ij} is a binary response corresponding to (i,j)th unit in the data with probability P_{ij} , $\mathbf{x}'_{ij}$ is a row of a design matrix; β_0 is an intercept, $\boldsymbol{\beta}$ is a vector of fixed effects, δ_i is a random effect with $\delta_i \sim N(0, \sigma^2_\delta)$ and ϵ_{ij} is an error term with $\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$.

If responses are ordered categorical with a factor having random effect on the responses, then a mixed cumulative logit model is more appropriate for modeling such data. The cumulative logit model is also called proportional odds model because of constant proportionality to each logit [Agresti, 2002]. For example, the mixed cumulative logistic model can be written as

$$\begin{aligned} Y_{ij} \mid \delta_i, \epsilon_{ij} &\sim \text{Multinomial}(1, \boldsymbol{P_{ij}}) \\ \text{logit}[P(Y_{ijk} > c \mid \mathbf{x})] = \log\left(\frac{\sum_{l=c+1}^{T} P_{ijl}}{\sum_{l=1}^{c} P_{ijl}}\right) &= \beta_{c0} + \mathbf{x}'_{ij}\boldsymbol{\beta} + \delta_i + \epsilon_{ij} \end{aligned} \tag{3.2}$$

where Y_{ij} is a ordered categorical response corresponding to (i,j)th unit with probability P_{ij}, $i = 1, 2, \ldots, b$, $j = 1, 2, \ldots, n_i$, b is the number of units in the higher stratum, n_i is the number of sub-units in the ith stratum; β_c is an intercept corresponding to response category c, $c = 1, 2, \ldots, T-1$; the $\{\beta_c\}$ is decreasing in c, because logit$[P(Y_{ijk} > c \mid \mathbf{x})]$ decreases in c for fixed $\mathbf{x}$ and logit is a decreasing function of $P(Y_{ijk} > c \mid \mathbf{x})$, β is a vector of fixed effects, δ_i is a random effect with $\delta_i \sim N(0, \sigma_\delta^2)$, ϵ_{ij} is an error term with $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. Further details on cumulative logit are available in Agresti [2002].

3.3 Model Selection

Model selection is a process of estimating the performance of different models in order to choose the best approximate one. We have competing models during Bayesian model building stages in the current study where the models are compared on the basis of Deviance Information Criterion (DIC). The DIC is based on trade-off between the fit of the data to the model and the corresponding complexity of the model [Spiegelhalter et al., 2002]. The DIC is defined as

$$\text{DIC} = \text{deviance} + \text{complexity}$$

The deviance is defined as $D(\theta) = -2 \log L(y|\theta)$ which quantifies the badness of fit; the complexity is measured by an estimate of the effective number of parameters $p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \overline{D} - D(\overline{\theta})$ i.e. posterior mean deviance minus deviance evaluated at the posterior mean of the parameters. Then the DIC is defined analogous to AIC as $DIC = D(\overline{\theta}) + 2p_D = \overline{D} + p_D$. Models with smaller DIC are better supported by the data.

We follow stepwise manual forward selection process to choose a model starting with no potential predictor variables. At each step, a predictor is added in such a way that the resulting model has the lowest DIC value. The predictors concerning fixed effects are considered first, then random effects are added in the model. The process is stopped when there is no further decrease in DIC values by adding or subtracting any predictor variables or their interactions or any random effects in the model.

3.4 The Polypropylene Industrial Experiment

3.4.1 What is Polypropylene?

Polypropylene is a type of thermoplastic polymer resin. It is used both in the average household and in the commercial and industrial applications. The chemical designation of polypropylene is C_3H_6. One of the benefits of using this type of plastic is that it can be useful in numerous applications including as a structural plastic or as a fiber-type plastic. The polypropylene experiment is of interest to car manufacturers due to the frequent use of propylene in the car industry. This is because polypropylene is inexpensive, light and environmentaly friendly [Jones and Goos, 2007].

3.4.2 Underlying Design

The current study is motivated by a polypropylene industrial experiment where responses are continuous and categorical. Four Belgian companies, namely Domo Polypropylene Compounds (a producer of thermoplastic materials), Europlasma (a developer of gas plasma systems), Structuplas (a company that specializes in the finishing of thermoplastic materials) and Techni-Coat International (a company that specializes in coatings) conducted the experiment to investigate the effect of several additives and a gas plasma treatment on the adhesive properties of polypropylene.

An undesirable property of polypropylene is that glues and coatings do not adhere well to its surface unless it undergoes a surface treatment, like a gas plasma treatment. It was a matter of investigation to look for economical plasma treatments which lead to good adhesion to various coatings. Four experimental factors related to the plasma treatment, each at three levels, are the gas flow rate, the power, the reaction time, and the type of gas used. Three types of gas, namely one etching gas and two activation gases, were used in the experiment. As decided by the plant engineers after some pilot tests, the gas flow rates used in the experiment were between 1000 and 2000 sccm, power ranged from 500 to 2000 W and reaction times lay between 2 and 15 min.

Table 3.1: Levels of factors studied in the polypropylene experiment

Factor	Units	Levels		
		-1	0	1
EPDM (X_1)	%	0		15
Ethylene (X_2)	%	0		10
Talcum (X_3)	%	0		20
Mica (X_4)	%	0		20
Lubricant (X_5)	%	0		1.5
UV-stabilizer (X_6)	%	0		0.8
EVA (X_7)	%	0		1.5
Power (X_8)	Watts	500	1000	2000
Gas flow rate (X_9)	sccm	1000	1500	2000
Processing time (X_{10})	min	2	8	15
Type of gas (X_{11})		Etching	Activation 1	Activation 2

The effects of several additives were studied in addition to the plasma treatments. Polypropylene is often compounded with additives such as stabilizers against ultraviolet (UV) light, lubricants, talcum, mica and/or colour pigments to tailor the plastic to a specific end use. It was strongly believed that some of the additives had an effect on the adhesive properties. Eventually, seven additives, each at two levels, namely ethylene diene monomer (EPDM) rubber, ethylene copolymer content of the rubber, talc, mica, lubricant, UV stabilizer, and ethylene vinyle acetate, were included in the study. The levels and units used for each of these eleven factors, coded as $X_1 - X_{11}$, are shown in Table 3.1. In the study, seven factors namely EPDM, ethylene, talcum, mica, lubricant, UV stabilizer, EVA are considered as whole plot factors which are hard to change, and four sub-plot factors are power, gas flow, processing time and type of gas.

The entire polypropylene experiment involves a complicated randomization which is due to the fact that the complete experiment was carried out in several stages. The sequence of events in polypropylene experiment is illustrated in Figure 3.1.

At the beginning, 20 batches of polypropylene plates were produced according to the whole-plot design for the seven additives. Each of the batches consists of several dozen polypropylene plates with the same settings for the seven additives. Each of the plates was stored in identical conditions. For the subsequent stages, the appropriate number of plates was removed from storage immediately prior to further processing.

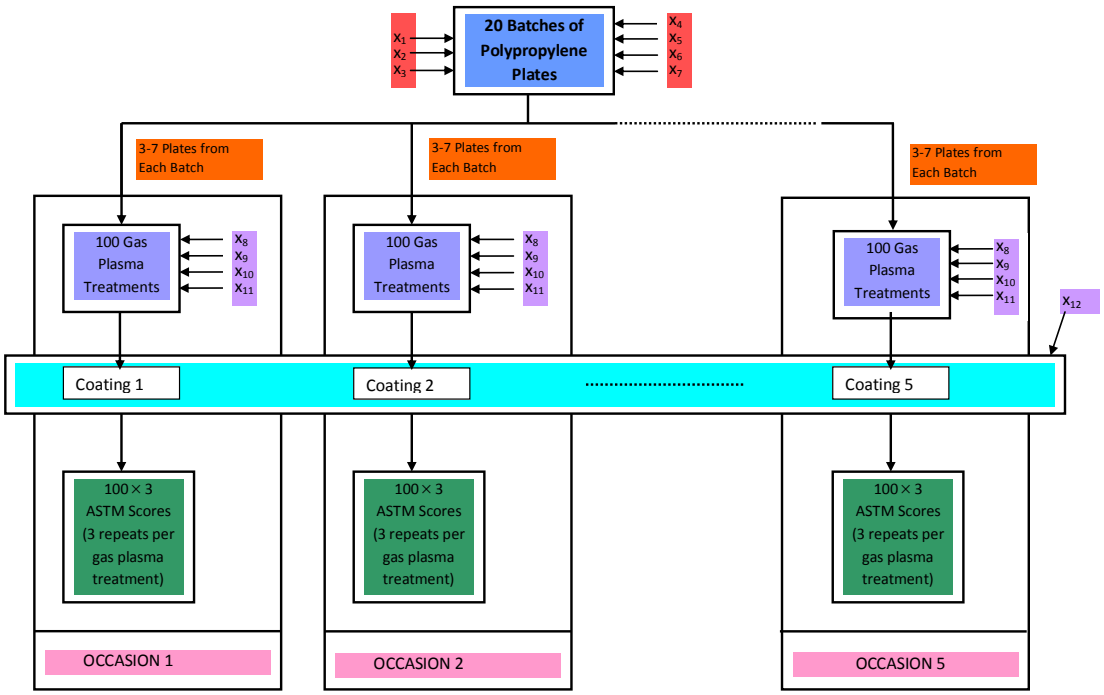

Figure 3.1: Factors and stages in the polypropylene experiment

In the next stage, three to seven sets of three samples were randomly selected from each of the 20 batches. The three samples in each set were processed together in one oven run, using one gas plasma treatment according to the sub-plot design. The sub-plot design consisted of 100 gas plasma treatments which were applied in 100 independent oven runs. A fixed number of days after the gas plasma treatment, coating 1 was applied to each of three samples in a set (three repeats). A six-level categorical response, related to the success of the coating's adhesion to the plastic, was measured as soon as the coating became dry. This stage was repeated four more times for other four different types of coatings. Thus, in the experiment, there were 100 runs each with 3 repeats, implies there were 300 measurements for each of the coatings and therefore, the whole experiment was supposed to contain 1500 measurements in total for all 5 coatings. However, due to having some missing values there were 1492 measurements in total and further details are available

elsewhere, Goos and Gilmour [2012] and Jones and Goos [2007].

Table 3.2: Frequency distribution of ASTM scores in the polypropylene experiment

ASTM	Coating 1	Coating 2	Coating 3	Coating 4	Coating 5
0	61	39	40	220	69
1	13	14	11	7	17
2	26	29	19	13	29
3	34	38	23	35	32
4	31	57	7	9	123
5	135	123	200	10	28
Total	300	300	300	294	298

The adhesion quality was assessed visually using the American Standard Test Method (ASTM) score which is considered standard for adhesion testing. The test method involves a six-point scale. A frequency distribution of the ASTM scores obtained in the polypropylene experiment is presented in Table 3.2. For coatings 1, 2, and 5 the outcome categories (ASTM scores) are approximately evenly distributed. For coating 3, two thirds of the measurements resulted in an ASTM score of 5, whereas for coating 4, more than 80% of the measurements resulted in an ASTM score of 0. Also, there are small numbers of missing observations for coatings 4 and 5.

3.5 Binary Response Data Analysis

In the test, a coating was considered acceptable if it resulted in an ASTM score of at least three. Thus, success of a coating was defined as

$$\text{Success of coating} = \begin{cases} 1 & \text{if} \quad \text{ASTM} \geq 3 \\ 0 & \text{if} \quad \text{ASTM} \leq 2 \end{cases}$$

The success of a coating is a binary response variable for which a mixed binary logit model is an appropriate choice. Following the model (3.1) the appropriate

model for binary response analysis would be

$$\begin{aligned} Y_{ijk} \mid \delta_i, \epsilon_{ij} &\sim \text{Bernoulli}(P_{ij}) \\ \text{logit}(P_{ij}) &= \beta_0 + \mathbf{x}'_{ij}\boldsymbol{\beta} + \delta_i + \epsilon_{ij} \end{aligned} \tag{3.3}$$

where Y_{ijk} is the response from the kth test, $k = 1, 2, 3$, on the jth oven run, $j = 1, 2, \ldots, n_i$, from the ith batch, $i = 1, 2, \ldots, 20$, P_{ij} is the probability of success for the jth oven run from batch i, β_0 is an intercept, $\mathbf{x}'_{ij}$ is the transpose of a row of the design matrix corresponding to ith batch and jth oven run, $\boldsymbol{\beta}$ is a vector of fixed factor effects, $\delta_i \sim N(0, \sigma^2_\delta)$ is a random effect due to batch i, and $\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$ is random effect due to jth oven run from batch i.

Goos and Gilmour [2012] analyzed the polypropylene data by likelihood-based methods using SAS procedure GLIMMIX. The GLIMMIX performs estimation and statistical inference for generalized linear models (GLMMSs) [Schabenberger, 2005]. Restricted pseudo-likelihood (RPL) is the default estimation method in PROC GLIMMIX for models containing random effects [Wolfinger and O'Connell, 1993]. However, the variance component estimates obtained by the GLIMMIX were not accurate due to having insufficient number of whole plots (batches) in the polypropylene experiment. The variance components due to batches (σ^2_δ) were estimated to be zero for coating 3 and coating 4 in binary response analyses and for coating 1, coating 4, and coating 5 in cumulative logit analyses. As the zero variance-components are unrealistic, a Bayesian analysis was recommended to avoid misleading inferences possibly arising in likelihood-based methods.

The unknown parameters in model (3.3) are β_0, $\boldsymbol{\beta}$, σ^2_δ and σ^2_ϵ for which we need to specify priors for these parameters in Bayesian approach. Since the main objective of the study is to estimate the fixed effects, it might be reasonable to use a non-informative, or a weakly informative prior for the parameters corresponding to fixed effects. We consider priors of the form $\beta_r \sim N(0, \tau_r)$ where β_r is the effect corresponding to rth factor or interaction and $\tau_r = 1/\sigma^2_r$). The prior for β_r can be non-informative by taking σ^2_r very large. For variance components, we need to consider priors carefully. The gamma family of priors is used for precision parameter i.e. for the variance components. In our data, we have only two strata batch and run, and the corresponding variance components are σ^2_δ and σ^2_ϵ. As σ^2_ϵ

will be well estimated with lots of degrees of freedom from the data its prior should be less influential, thus we might use a non-informative prior. It is often easier to express the prior information about intra-main plot correlation $\rho = \frac{\sigma_\delta^2}{\sigma_\delta^2+\sigma_\epsilon^2}$ or equivalently $\rho = \frac{\tau_\epsilon}{\tau_\epsilon+\tau_\delta}$, where $\tau_\epsilon = 1/\sigma_\epsilon^2$, $\tau_\delta = 1/\sigma_\delta^2$. Often a beta distribution is appropriate for the single parameter ρ. We have analyzed each of the five coatings by Bayesian methods using WinBUGS 1.4.

3.5.1 Binary Response Analysis of Coating 1

We assume the priors concerning model (3.3) to analyze coating 1 in Bayesian method. We consider slightly informative priors for factor coefficients and a modest prior for batch effects, for instance, $\beta_i \sim N(0, 0.01)$, $\tau_\epsilon \sim \text{Gamma}(1, 1)$, and $\rho \sim \text{beta}(5, 5)$, where $\tau_\epsilon = 1/\sigma_\epsilon^2$, $\tau_\delta = 1/\sigma_\delta^2$ and $\rho = \frac{\tau_\epsilon}{\tau_\epsilon+\tau_\delta}$. The reason behind considering this slightly informative prior for factor effects is that there arises convergence problem if we use non-informative priors e.g. $\beta_i \sim N(0, 0.0001)$ instead of $\beta_i \sim N(0, 0.01)$ during binary response analysis.

For the binary response data, we have obtained the Bayesian results for coating 1 displayed in Table 3.3 following the model simplified by Goos and Gilmour [2012] who used a forward selection procedure, where a main-effects model was estimated first and interaction effects among factors with significant main effects were added one by one. We did not attempt to search the best models independently in Bayesian approach because of severe convergence problems. At the beginning we started with full model containing all main effects and two factor interactions and wanted to implement backward elimination method. Due to convergence problem, it was impossible to implement this technique. Then we started method of forward selection starting with a single factor and gradually adding other factors and their interactions in the model. In the selection process, models were compared with the DIC values obtained at the end of fitting a model. A model with minimum DIC ruled out a competing model with larger DIC value. During this process, while adding some of the factors and interactions it becomes impossible to get estimates due to convergence problem and therefore, left the idea of searching models independently in the Bayesian method. Thus, to compare Bayesian and likelihood estimates, we were satisfied with the simplified models rather than ending up with wrong models describing as the best. However, estimating the mixed binary logit

Table 3.3: Classical and Bayesian estimates obtained from mixed binary logit model for coating 1

	Classical Method		Bayesian Method		
Effect	Estimate	P-value	Mean	SD	(95% CI)
Intercept	1.768		2.783		
EPDM	0.989	0.011	1.573	0.669	(0.299 , 2.874)
Ethylene	0.765	0.034	1.330	0.661	(0.141 , 2.829)
Talcum	0.978	0.015	1.528	0.688	(0.271 , 2.962)
Time	1.473	0.000	2.256	0.511	(1.381 , 3.404)
Gas type	1.763	0.000	2.790	0.751	(1.397 , 4.449)
Batch (σ^2_δ)	0.517		4.387	2.976	(0.987 , 12.310)
Run (σ^2_ϵ)	2.295		6.188	3.014	(2.374 , 14.040)

model turned out to be a real challenge for Goos and Gilmour [2012] also when they attempted to include certain interactions in the model. For instance, adding the two factor interaction effect of EPDM and type of gas caused the convergence to fail in their study.

The posterior estimates of coating 1 are displayed in Table 3.3 along with the likelihood-based estimates that was obtained by Goos and Gilmour [2012]. In Bayesian estimation regarding coating 1 sample size was 9000, highest feasible number for coating 1, thinning was 10 to avoid autocorrelation effects in the posterior estimates. A burn-in of 1000 was allowed and then samples of 800 were used to calculate the estimates. The reason behind this unusual sample size is a convergence problem. Perhaps, a sample with size 800 is reasonable to compute posterior estimates as kernel densities relevant to the parameters were smooth and unimodal. With the effort of taking large samples the WinBUGS programme stuck at some points and failed to produce posterior estimates. Therefore, the sample size was kept moderate across the coatings during Bayesian estimations of mixed binary logit analysis. However, it is essential to give explanation why WinBUGS software failed in the current situation. We know that Gibbs sampling works iteratively by drawing samples from the full conditional distributions of unobserved nodes. From empirical studies, it is found that the Gibbs sampler cannot draw sample from non-informative priors in particular situations, for instance, in case of dealing with mixed models associated with binary responses and non-informative priors. While convergence problem arises during analysis of mixed binary logit models, WinBUGS provides error messages by labeling "Trap 66" which still re-

Table 3.4: Classical and Bayesian estimates obtained from mixed binary logit model for coating 2

	Classical Method			**Bayesian Method**		
Effect	Estimate	SE	P-value	Mean	SD	95% CI
Intercept	4.159			4.237		
EPDM	1.196	0.602	0.075	1.240	0.637	(0.048, 2.558)
Ethylene	1.569	0.588	0.023	1.752	0.610	(0.674, 3.046)
Talcum	2.179	0.744	0.013	2.111	0.706	(0.730, 3.493)
Mica	1.332	0.684	0.083	0.941	0.562	(-0.181, 2.023)
Power	0.805	0.374	0.036	-0.735	0.390	(-1.512, 0.025)
Time	2.636	0.501	0.000	2.803	0.557	(1.798, 3.964)
Type of gas	2.987	0.689	0.000	2.341	0.567	(1.291, 3.480)
Act. gas	-1.445	0.379	0.000	-1.743	0.465	(-2.768, -0.928)
Power×Act. gas	1.365	0.422	0.002	1.356	0.442	(0.497, 2.271)
Batch (σ_δ^2)	3.667			3.813	2.495	(0.915, 10.180)
Run (σ_ϵ^2)	1.115			3.390	2.109	(0.826, 8.874)

mains as a "black box" to the applied Bayesian practitioners. Perhaps for deeper understanding of the problem in the current situation, it might be essential to work closely with the WinBUGS developer team.

We find in classical method that the main effects of EPDM, ethylene, talcum, time and gas type (etching gas versus activation gas) are significant at 5% level for coating 1 shown in Table 3.3. All the factors just mentioned also appeared to be important in Bayesian method as all of the estimates were within 95% central Bayesian credible intervals. The estimates of the variance components in classical methods were $\hat{\sigma}_\delta^2 = 0.5170$, and $\hat{\sigma}_\epsilon^2 = 2.2950$ which are differing with the corresponding Bayesian estimates. This could be due to the priors assumed for the parameters.

Both likelihood and Bayesian approaches show that the estimates of σ_δ^2 are positive which is an indication of having batch to batch variability. On the other hand, the estimates of σ_ϵ^2 confirms that the three observational units (i.e. tests) within every experimental unit (i.e. runs) are dependent.

3.5.2 Binary Response Analysis of Coating 2

We find that the main effects of ethylene, talcum, power and time are significant at 5% level in Table 3.4. The type of gas, the type of activation gas (activation gas 1 versus 2) and an interaction terms are also significant. However, the effects of EPDM and mica are borderline significant.

For coating 2, we assumed slightly informative priors for factor coefficients $\beta_r \sim N(0, 0.01)$ and a modest prior for batch effects $\tau_\epsilon \sim$ Gamma$(1, 1)$, and $\rho \sim$ beta$(5, 5)$, where $\tau_\epsilon = 1/\sigma_\epsilon^2$, $\tau_\delta = 1/\sigma_\delta^2$ and $\rho = \frac{\tau_\epsilon}{\tau_\epsilon + \tau_\delta}$. For the binary response data we have obtained the posterior estimates displayed in Table 3.4 following the model simplified by Goos and Gilmour [2012]. The sample size was 11000, thinning was 20 to avoid autocorrelation effects in the posterior estimates. A burn-in of 1000 was allowed and then samples of 500 were used to calculate the estimates. The estimates of the variance components in classical methods were $\hat{\sigma}_\delta^2 = 3.667$, and $\hat{\sigma}_\epsilon^2 = 1.1154$ which are slightly differing with the corresponding Bayesian estimates shown in Table 3.4. This could be due to the priors assumed for the parameters.

Both likelihood and Bayesian methods produce similar output for coating 2. Table 3.4 shows that factors effects did not include 0 in their 95% credible intervals except mica and power. However, the effect of power on responses appeared to have the reverse sign in Bayesian method for coating 2. Perhaps power effect is negligible for coating 2 responses as in Bayesian method the estimate corresponding to power contains zero in its 95% Bayesian credible interval. Both likelihood and Bayesian approaches show that the estimates of σ_δ^2 are positive which is an indication of having batch to batch variability. On the other hand, the estimates of σ_ϵ^2 confirms that the three observational units (i.e. tests) within every experimental unit (i.e. runs) are dependent.

3.5.3 Binary Response Analysis of Coating 3

The posterior results obtained from binary analysis of coating 3 following the simplified model by Goos and Gilmour [2012] are presented in Table 3.5. The main effects EPDM, ethylene, time and type of gas appeared to be important in Bayesian

Table 3.5: Classical and Bayesian estimates obtained from mixed binary logit model for coating 3

	Classical Method		Bayesian Method		
Effect	Estimate	P-value	Mean	SD	(95% CI)
Intercept	4.268		4.811		
EPDM	1.773	0.000	2.055	0.574	(1.021 , 3.256)
Ethylene	1.666	0.001	1.896	0.587	(0.765 , 3.195)
Talcum	0.928	0.055	0.788	0.602	(-0.442 , 1.969)
Mica	0.812	0.100	0.658	0.550	(-0.444 , 1.703)
Lubricant	0.869	0.038	0.790	0.502	(-0.234 , 1.795)
Time	2.601	0.000	2.556	0.503	(1.662 , 3.563)
Gas type	3.365	0.002	3.714	1.023	(1.900 , 5.956)
Batch (σ^2_δ)	0.000		1.521	1.855	(0.033 , 7.358)
Run (σ^2_ϵ)	3.782		6.402	3.342	(1.947 , 14.430)

analysis as all of the effects do not include 0 in their 95% credible intervals, though in classical method lubricant was significant at 5% level and talcum and mica were borderline significant, they were found not to be important in Bayesian analysis. The batch to batch variation was estimated to be 0 for coating 3 in classical method, however, it was non-zero in Bayesian analysis.

3.5.4 Binary Response Analysis of Coating 4

The coating 4 had troublesome behaviour as we faced severe convergence problem during the binary response analysis. Perhaps the reason is that we have more than three quarters of the measurements of coating 4 resulting in an ASTM score of 0. The results of coating 4 from mixed binary logit are presented in Table 3.6. Only the effect gas type was significant and time was borderline significant in classical analysis. Though the variance component estimates were either 0 or moderate in classical analysis, however, in Bayesian analysis these point estimates were unreasonably high. As the credible intervals of these estimates are also wide, the main problem might not be in the estimation method rather than data of coating 4 itself.

3.5.5 Binary Response Analysis of Coating 5

The results obtained from binary analysis of coating 5 are given in Table 3.7. The main effects EPDM, talcum, mica, power, type of gas and interactions Power×Gas

Table 3.6: Classical and Bayesian estimates obtained from mixed binary logit model for coating 4

	Classical Method		**Bayesian Method**		
Effect	Estimate	P-value	Mean	SD	(95% CI)
Intercept	-2.108		-2.473		
Time	0.690	0.052	1.064	0.658	(-0.244 , 2.331)
Gas type	1.186	0.005	1.402	0.692	(0.172 , 2.864)
Batch (σ^2_δ)	0.000		17.300	15.920	(2.236 , 59.310)
Run (σ^2_ϵ)	3.778		22.800	14.350	(5.645 , 58.820)

Table 3.7: Classical and Bayesian estimates obtained from mixed binary logit model for coating 5

	Classical Method		**Bayesian Method**		
Effect	Estimate	P-value	Mean	SD	(95% CI)
Intercept	2.264		2.784		
EPDM	1.293	0.004	1.762	0.527	(0.824 , 2.917)
Ethylene	0.724	0.061	1.108	0.488	(0.248 , 2.174)
Talcum	1.985	0.001	2.634	0.644	(1.476 , 4.027)
Mica	1.350	0.009	1.596	0.572	(0.484 , 2.750)
Power	1.068	0.001	1.232	0.363	(0.556 , 1.972)
Time	1.559	0.307	1.888	0.401	(1.165 , 2.752)
Gas type	1.658	0.001	1.862	0.507	(0.949 , 2.904)
Act. Gas	-0.098	0.670	-0.210	0.295	(-0.790 , 0.349)
Power × Gas type	1.765	0.003	1.864	0.632	(0.715 , 3.167)
Power × Act. Gas	0.499	0.076	0.639	0.368	(-0.034 , 1.426)
Time × Gas type	0.935	0.041	0.945	0.548	(-0.096 , 2.047)
Batch (σ^2_δ)	1.349		2.303	1.511	(0.572 , 6.445)
Run (σ^2_ϵ)	0.542		1.839	1.075	(0.493 , 4.738)

type, Time×Gas type were significant at 5% level and ethylene, time, activation gas and the interaction Power×Act. gas were borderline significant in likelihood-based analysis. All the factors and an interaction but Time×Gas type and Power×Act. gas were found to be important in the Bayesian analysis.

3.5.6 Remarks on Mixed Binary Logit Analysis of Coatings

With classical methods, analyzing the data for coating 3 and 4 led to several differences with the results obtained for coatings 1, 2, and 5. For both coatings 3 and 4, σ^2_δ were estimated to be zero which implied that there was no batch-to-batch

variation. Actually zero variability of a variance component is unrealistic. The Bayesian method ensures that none of the estimates of σ_δ^2 can be zero. The batch-to-batch variation was similar for all coatings except coating 4. However, perhaps the batch-to-batch variability could not be measured properly for coating 4 due to convergence problems in the mixed binary analysis. The variability due to runs were more or less similar for all coatings except for coating 4. The estimates of fixed effects for all coatings were similar to the estimates in classical methods both in magnitude and direction except for the coefficient of power which is reversed in sign.

The analysis of binary data frequently presents problems for which there are no standard solutions, as pointed out by Collett and Stepniewska [1999]. This appeared to be true in our whole binary data analysis. Identification of subsets of significant explanatory variables and their interactions was difficult for all coatings, as we encountered many problems similar to Goos and Gilmour [2012] in fitting mixed models for binary data. This is not unusual to have severe problems with fitting GLM for categorical data as mentioned by Chipman and Hamada [1996] also. Perhaps these are due to separation problems arising in binary regression model [Goos and Gilmour, 2012]. The main aim of this study was to estimate the main effects of each of the factors as well as all relevant interactions in Bayesian method overcoming the limitations of classical methods and thereby to compare two approaches. However, for binary data possibly it was an ambitious goal because of convergence problems while including several interactions. We followed manual forward selection, where we fit a model with main effects first, and then added interactions involving important main effects one by one. Slightly informative priors for factor coefficients and more accurate set of initial values often improve convergence problems in mixed binary logit model analysis.

3.6 Ordinal Response Data Analysis

We expect that analyzing actual ASTM scores would provide more information than binary analyses. The cumulative logit model [Agresti, 2002] is a better tool to analyze naturally ordered ASTM scores. A suitable model for ASTM scores

seems to be

$$\begin{aligned} Y_{ijk} \mid \delta_i, \epsilon_{ij} &\sim \text{Multinomial}(1, \boldsymbol{P_{ij}}) \\ \text{logit}[P(Y_{ijk} > c)] &= \log\left(\frac{\sum_{l=c+1}^{6} P_{ijl}}{\sum_{l=1}^{c} P_{ijl}}\right) = \beta_{c0} + \mathbf{x}'_{ij}\boldsymbol{\beta} + \delta_i + \epsilon_{ij} \end{aligned} \tag{3.4}$$

where Y_{ijk} is the response from kth test on the jth run from the ith batch, P_{ijl} is the probability for the ASTM score l, in the jth oven run from batch i, $\mathbf{P}'_{ij} = [P_{ij1}, P_{ij2}, \ldots, P_{ij6},]$, β_{c0} is an intercept corresponding to response category c, $\mathbf{x}_{ij}$ is a design matrix corresponding to ith batch and jth oven run, $\boldsymbol{\beta}$ is a vector of fixed effects due to factors, $\delta_i \sim N(0, \sigma^2_\delta)$ is a random effect due to batch i, $\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$ is random effect due to jth oven run from batch i, and all random variables are independent. The intercepts are the only parameters which depend on the comparison of categories being made. The model has the same effects $\boldsymbol{\beta}$ for each logit. Thus, we assumed the proportionality of odds in the analysis. However, it is difficult to test the deviation of assumption of proportionality in such a complex model in classical approach, and the same is true for the Bayesian analyses.

The ordered ASTM scores were analyzed for the five coatings using general model (3.4) in classical methods by Goos and Gilmour [2012] and in Bayesian methods in this study. The following sections will describe Bayesian investigation of all the coatings.

3.6.1 Ordinal Response Analysis of Coating 1

The unknown parameters in model (3.4) are β_{c0}, $\boldsymbol{\beta}$, σ^2_δ, and σ^2_ϵ. Priors were assumed for intercepts (β_{c0}), fixed effects of factors ($\boldsymbol{\beta}$), and for the elements of random effects. We used normal priors for fixed effects ($\boldsymbol{\beta}$) and ordered normal priors for β_{c0}. The priors for the intercept β_{c0} is ordered normal because logits are either increasing or decreasing with the ordered categories of responses. Informative priors were assumed for process variation σ^2_δ but non-informative or vague priors for factor parameters as no substantial information was available for them. To analyze data of coating 1 the priors assumed corresponding to model (3.4) are $\beta_{c0} \sim \text{N}(0, 0.0001)$, $\beta_r \sim \text{N}(0, 0.0001)$, $\tau_\epsilon \sim \text{Gamma}(1, 1)$, $\rho \sim \text{Beta}(5, 5)$, where $\rho = \frac{\tau_\epsilon}{\tau_\epsilon + \tau_\delta}$, $\tau_\epsilon = 1/\sigma^2_\epsilon$, and $\tau_\delta = 1/\sigma^2_\delta$.

We did independent search of Bayesian models applying manual forward selection method and the models were compared on the basis of DIC values for coating 1. However, we ended up with the same model as obtained by Goos and Gilmour [2012] for coating 1. The results from the likelihood estimation and the posteriors obtained from the fitted Bayesian model are presented in Table 3.8. In classical method, the main effects of the factors EPDM, ethylene, talcum, mica, time, gas type and activation gas are clearly highly significant. Also, the interaction of EPDM and ethylene appeared to be significant. However, Time^2, EPDM$\times$Act. gas are borderline significant. In the Bayesian analysis the factors EPDM, ethylene, talcum, and time are remarkably important as 95% Bayesian credible intervals include the estimates of these factor effects. However, mica and all the interactions were found not to be important in Bayesian analysis. In classical method, the batch-to-batch variation was inestimable, on the other hand, Bayesian techniques provide the estimate of batch-to-batch variation (σ^2_δ) as 3.6190. In classical and Bayesian methods the variance component due to runs (σ^2_ϵ) were 3.8584 and 6.8210 respectively.

Table 3.8: Classical and Bayesian estimates obtained from mixed cumulative logit model for coating 1

	Classical Method		**Bayesian Method**		
Effect	Estimate	P-value	Mean	SD	95% CI
Intercept 1	4.947		6.396		
Intercept 2	4.283		5.602		
Intercept 3	3.188		4.317		
Intercept 4	1.945		2.867		
Intercept 5	0.825		1.568		
EPDM	0.918	0.001	1.235	0.575	(0.143, 2.413)
Ethylene	0.853	0.001	1.224	0.569	(0.143, 2.385)
Talcum	1.116	0.001	1.590	0.671	(0.336, 2.961)
Mica	0.751	0.019	1.280	0.718	(-0.075, 2.750)
Time	1.788	0.000	2.629	0.462	(1.809, 3.602)
Gas Type	2.102	<.000	2.995	0.598	(1.899, 4.229)
Act. gas	-0.778	0.009	-1.149	0.403	(-1.973, -0.386)
EPDM$\times$Ethylene	0.752	0.004	1.074	0.556	(-0.016, 2.196)
Time^2	-1.153	0.057	-1.494	0.844	(-3.193, 0.124)
EPDM$\times$Act.gas	0.549	0.061	0.732	0.394	(-0.017, 1.522)
Batch (σ^2_δ)	0.000		3.619	2.203	(0.899 , 9.157)
Run (σ^2_ϵ)	3.858		6.821	2.252	(3.482 , 12.340)

3.6.2 Ordinal Response Analysis for Coating 2

We assume the priors corresponding to model (3.4) with a view to analyzing coating 2 as $\beta_{c0} \sim \mathrm{N}(0, 0.0001)$, $\beta_r \sim \mathrm{N}(0, 0.0001)$, $\tau_\epsilon \sim \mathrm{Gamma}(1, 1)$, $\rho \sim \mathrm{Beta}(5, 5)$, where $\rho = \frac{\tau_\epsilon}{\tau_\epsilon + \tau_\delta}$, $\tau_\epsilon = 1/\sigma_\epsilon^2$, and $\tau_\delta = 1/\sigma_\delta^2$.

The likelihood estimates of coating 2 by Goos and Gilmour [2012] and the Bayesian estimates are presented in Table 3.9. The main effects of the factors EPDM, ethylene, talcum and time are clearly highly significant. Also, the type of gas (etching gas versus activation gas) and type of activation gas (activation gas 1 versus 2) have significant effects. The factor power does not have a significant main effect, but its interactions with the type of gas and the type of activation gas are highly significant. There is also some indication that the interaction between mica and the type of activation gas has an effect on the ASTM score. In likelihood-based analysis, nonsignificant main effects talcum and mica were included in the model due to marginality restriction. The principle of marginality refers to the fact that the main effects should be included in the model when one or more interactions involving them found to be important even though individual effects are negligible.

Mixed cumulative logit analysis detects more effects than mixed binary logit analysis. Comparing Table 3.4 and Table 3.9 we find the effects EPDM×Ethylene, EPDM×Talcum , Power×Gas type, Power×Time, Ethylene×Power, Mica×Act.gas were not detected in binary analysis under likelihood method. Further, we find more important effects using mixed cumulative logit model for ASTM scores with the Bayesian method than with the likelihood-based method for coating 2 shown in Table 3.9. In Bayesian approach the newly identified factors that are influencing ASTM scores are EPDM×Ethylene, EPDM×Talcum, and Power×Time for coating 2.

The two variance components in the model σ_δ^2 and σ_ϵ^2 are estimated to be 1.2507 and 3.7213 respectively in the likelihood method while their counterparts in Bayesian methods are 4.1310 and 5.0500 respectively. The variance components being positive in both methods suggest the existence of substantial batch-to-batch variation as well as dependence between the three observational units (i.e. tests) within every experimental unit (i.e. within every oven run).

Table 3.9: Classical and Bayesian estimates obtained from mixed cumulative logit model for coating 2

	Classical Method		Bayesian Method		
Effect	Estimate	P-value	Mean	SD	95% CI
Intercept 1	6.209		7.546		
Intercept 2	4.916		5.920		
Intercept 3	3.117		3.740		
Intercept 4	1.404		1.698		
Intercept 5	-0.499		-0.478		
EPDM	0.741	0.065	0.803	0.578	(-0.314 , 1.996)
Ethylene	1.315	0.003	2.063	0.590	(0.988 , 3.310)
Talcum	1.487	0.005	2.083	0.697	(0.766 , 3.510)
Mica	0.757	0.121	1.199	0.706	(-0.137 , 2.633)
Power	-0.313	0.295	-0.340	0.352	(-1.033 , 0.345)
Time	1.931	0.000	2.585	0.410	(1.829 , 3.437)
Gas type	2.383	0.000	3.321	0.567	(2.303 , 4.499)
Act. gas	-0.593	0.060	-0.775	0.362	(-1.516 , -0.080)
EPDM×Ethylene			-0.702	0.563	(-1.860 , 0.381)
EPDM×Talcum			-1.049	0.583	(-2.254 , 0.100)
Power×Gas type	1.212	0.013	1.853	0.582	(0.742 , 3.059)
Power×Act. gas	0.843	0.014	0.970	0.401	(0.178 , 1.770)
Power×Time			1.334	0.442	(0.506 , 2.258)
Ethylene × Power	-0.576	0.051	-0.644	0.351	(-1.357 , 0.044)
Mica × Act. gas	0.573	0.065	0.744	0.354	(0.068 , 1.448)
Batch (σ^2_δ)	1.251		4.131	2.216	(1.313 , 9.563)
Run (σ^2_ϵ)	3.721		5.050	1.707	(2.497 , 9.073)

3.6.3 Ordinal Response Analysis of Coating 3

The classical and Bayesian results of cumulative logit analysis of coating 3 are presented in Table 3.10. The factors EPDM, ethylene, time, gas type are highly significant in classical method. Though power is not significant, the interaction term Power×Gas type is significant at 5% level. In the selected Bayesian model EPDM, ethylene, time, gas type, activation gas found to be important. On the other hand, the factors talcum, lubricant and the interaction terms Talcum×Time and Lubricant×Power could be slightly important as 95% credible interval of mean effects corresponding to these factors include 0 marginally. The factor activation gas was not significant in classical method, however, it was not negligible in Bayesian method.

Table 3.10: Classical and Bayesian estimates obtained from mixed cumulative logit model for coating 3

	Classical Method		Bayesian Method		
Effect	Estimate	P-value	Mean	SD	(95% CI)
Intercept 1	6.418		10.450		
Intercept 2	5.490		9.252		
Intercept 3	4.179		7.546		
Intercept 4	2.745		5.714		
Intercept 5	2.328		5.159		
EPDM	1.622	0.008	3.252	1.022	(1.392 , 5.496)
Ethylene	2.065	0.002	3.795	1.078	(1.954 , 6.256)
Talcum			1.661	1.001	(-0.296 , 3.710)
Lubricant			1.642	0.941	(-0.164 , 3.624)
Power	0.392	0.439			
Time	2.8145	0.000	5.602	1.099	(3.693 , 8.006)
Gas Type	3.550	0.000	5.254	1.285	(3.012 , 8.067)
Act. gas	-0.887	0.175	-1.615	0.643	(-2.963 , -0.450)
Talcum×Time			1.506	0.873	(-0.133 , 3.345)
Power×Gas Type	1.822	0.046			
Lubricant×Power			-0.890	0.670	(-2.221 , 0.4208)
Batch (σ^2_δ)	2.452		8.288	5.438	(2.011 , 22.250)
Run (σ^2_ϵ)	4.272		12.780	5.439	(5.388 , 26.400)

3.6.4 Ordinal Response Analysis of Coating 4

The results obtained from cumulative logit analysis of coating 4 are presented in Table 3.11. The main effects time and gas type were significant at 5% level in classical method and these were reflected in Bayesian estimates as well by being within the 95% credible intervals. On the other hand, talcum was not significant at 5% level which was reflected in Bayesian analysis by including 0 in 95% credible interval. However, the results of coating 4 unusual as variance components due to batch ($\hat{\sigma}^2_\delta$) is zero, while the Bayesian counterpart is 55.76 and the component due to runs is 114.8 which seems to be quite unusual.

3.6.5 Ordinal Response Analysis of Coating 5

The classical and Bayesian estimates from cululative logit analysis of coating 5 are displayed in Table 3.12. The factors EPDM, ethylene, talcum, mica, lubricant, UV, power, time, gas type are highly significant in likelihood analysis. Though

Table 3.11: Classical and Bayesian estimates obtained from mixed cumulative logit model for coating 4

	Classical Method		**Bayesian Method**		
Effect	Estimate	P-value	Mean	SD	(95% CI)
Intercept 1	-1.240		-7.861		
Intercept 2	-1.906		-9.725		
Intercept 3	-3.011		-11.950		
Intercept 4	-5.884		-16.900		
Intercept 5	-7.371		-19.460		
Talcum	0.720	0.096	3.005	2.379	(-1.370 , 7.888)
Time	1.105	0.040	3.637	1.772	(0.385 , 7.345)
Gas type	1.521	0.012	5.001	2.108	(1.170 , 9.433)
Batch (σ^2_δ)	0.000		55.760	37.430	(12.810 , 154.100)
Run (σ^2_ϵ)	11.153		114.800	50.390	(42.130 , 243.900)

activation gas is not significant, however, its interaction with power appeared to be significant at 5% level. In Bayesian analysis the factors EPDM, ethylene, mica, lubricant, UV, power, time, gas type are found to be important as none of the credible intervals of these estimates include zero in their 95% credible intervals. The interactions 'Power×Gas type', 'Power×Act. gas', 'EPDM×Ethylene' found to be important both in classical and Bayesian methods. Though the estimate of variance component due to batch ($\hat{\sigma}^2_\delta$) was zero in likelihood-based method, it was estimated as 2.047 in Bayesian method.

3.6.6 Remarks on Mixed Cumulative Logit Analysis of Coatings

The results from the cumulative logit analysis of ASTM scores by classical and Bayesian methods for all coatings are displayed in Table 3.8, 3.9 , 3.10, 3.11, and 3.12. The factor EPDM has positive impact on four out of five coatings. The type of gas appeared to be important for all coatings. This was also confirmed in earlier classical and Bayesian analyses of binary response data. Some of the interactions appeared not be important in Bayesian analysis, for instance, Power×Type of Gas for coating 3. On the contrary some of the interactions were not detected in likelihood methods, for example, Power×Time, EPDM×Ethylene, EPDM×Talcum for coating 2.

Table 3.12: Classical and Bayesian estimates obtained from mixed cumulative logit model for coating 5

	Classical Method		Bayesian Method		
Effect	Estimate	P-value	Mean	SD	95% CI
Intercept1	5.090		5.395		
Intercept2	4.206		4.400		
Intercept3	3.039		3.101		
Intercept4	1.918		1.845		
Intercept5	-2.705		-4.024		
EPDM	1.236	0.000	0.588	0.451	(0.734, 2.543)
Ethylene	0.886	0.000	1.087	0.437	(0.226, 2.006)
Talcum	1.721	0.000	2.194	0.555	(1.125, 3.321)
Mica	1.146	0.000	1.513	0.552	(0.462, 2.631)
Lubricant	0.753	0.001	0.988	0.431	(0.147, 1.904)
UV	-0.924	0.000	-1.131	0.427	(-2.024, -0.331)
Power	0.923	0.001	1.214	0.331	(0.576, 1.895)
Time	1.612	0.000	2.120	0.354	(1.464, 2.818)
Gas Type	1.641	0.000	0.920	0.437	(1.117, 2.855)
Act. gas	-0.010	0.970	-0.093	0.323	(-0.729, 0.529)
Power×Gas Type	1.137	0.005	1.466	0.506	(0.500, 2.474)
Power×Act. Gas	0.715	0.019	0.929	0.380	(0.191, 1.701)
EPDM×Ethylene	0.537	0.019	-0.836	0.440	(-1.730, -0.027)
Power^2	-1.300	0.028	-0.991	0.639	(-2.246, 0.232)
Batch (σ^2_δ)	0.000		2.047	1.369	(0.457, 5.432)
Run (σ^2_ϵ)	2.753		4.	1.459	(2.179, 7.887)

One of the objectives of this Bayesian study was to obtain more reasonable estimates of the variance components. In classical method, the estimates of σ^2_δ were zero for some coatings. However, the variance components measured in Bayesian techniques are positive and are often more inflated than those of classical counterparts. One of the reasons could be due to the assumed priors for variance components in Bayesian approach. Perhaps assumed gamma prior is not suitable for the current study. Gelman [2006] identified serious problems with inverse-gamma family. Instead of gamma priors he suggested to use a uniform prior on the hierarchical standard deviation, using the half-t family when the number of groups is small and in other settings where a weakly informative prior is desired. In case of having multiple variance parameters in hierarchical modeling the half-t family was recommended. For the variance component corresponding to runs we used Gamma(1, 1) prior. Though the number of groups is not small, but due to the existence of multiple variance parameters the use of half-t family is justified. Later we will try

this half-t family during investigation of variance components with different priors in Section 3.8.

The models were selected on the basis of the deviance information criterion (DIC), a widely used criterion in Bayesian paradigm, which was discussed briefly in Section 3.3. The selected best model which produces Table 3.9 for coating 2 had the minimum DIC value 603.92 in comparison to other models.

Remarkably in the cumulative logit analyses under Bayesian method, we encountered no convergence problems except for coating 4. The unusual output for coating 4 also support this conclusion. The output is unusual in a sense that Bayesian estimates are widely differing with respect to classical estimates for coating 4 shown in Table 3.11. Also, all the intercepts are negative both in likelihood and Bayesian methods. The reason could be the failure to select the best model independently due to convergence problem in coating 4.

3.7 Combined Analysis of Coatings

One of the main goals of the polypropylene experiment was to compare the factor effects across the different coatings. This leads us to a combined analysis of all five coatings. Previously, we analyzed the data for each coatings separately in which case there are 11 factors and now the coating will act as a twelfth experimental factor. Though there was no formal randomization of coatings to occasions, the sequence in which they were run can be considered as being essentially random.

A Hasse diagram shown in Figure 3.2 can be used to understand and visualize the unit structure of multistratum design concerning the combined analysis.

A Hasse diagram is a simple graph, with nodes representing blocking factors and edges representing nesting relationships between blocking factors. The rules for constructing Hasse diagrams are described in Bailey [2008]. The Hasse diagram have two numbers adjacent to every node: the number of levels of the corresponding blocking factor and the corresponding degrees of freedom (in brackets). The

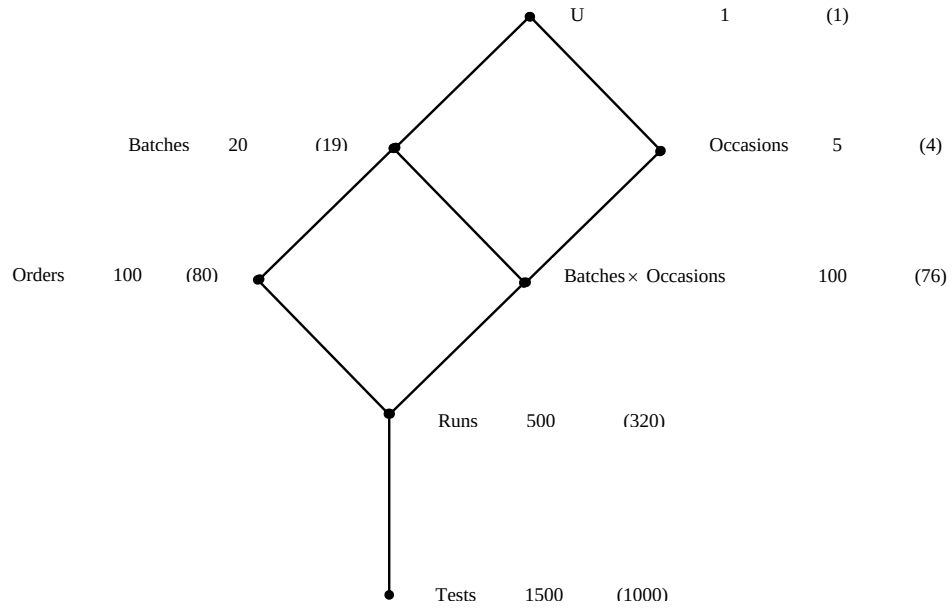

Figure 3.2: Hasse diagram for combined analysis

degrees of freedom corresponding to a node is obtained by subtracting the degrees of freedom for higher factors from the number of levels of the factor under consideration. In Figure 3.2, U denotes the universe that represents a fixed intercept parameter. The node U stands at the top of all nodes. Each node in the Hasse diagram represents a *stratum* in the analysis. The lowest stratum generally represents the observational units. In the combined analysis, there are 1500 tests, which belongs to the lowest stratum 'Tests', are the observational units shown in the Figure 3.2.

In the individual coating analyses, the random effects were due to batches and runs. The 100 oven runs were performed in the same order on each of the five occasions, which means that orders and occasions are crossed in the combined analysis. The combinations of orders and occasions define a stratum which correspond to the 500 oven runs in the Hasse diagram. Thus, there appears a random effect due to combinations of batches and occasions. Also, another random effect exists due to the stratum 'order' within batches.

Considering four random effects in the combined analysis of coatings, an appropriate model, as in Goos and Gilmour [2012], is

$$\begin{aligned} Y_{ijks} \mid \delta_i, \gamma_{ij}, \lambda_{ik}, \epsilon_{ijk} &\sim \text{Multinomial } (1, \boldsymbol{P_{ijk}}) \qquad (3.5) \\ \text{logit}\left[P(Y_{ijks} > c)\right] = \log\left(\frac{\sum_{l=c+1}^{6} P_{ijkl}}{\sum_{l=1}^{c} P_{ijkl}}\right) &= \beta_{c0} + \mathbf{x}'_{ijk}\boldsymbol{\beta} + \delta_i + \gamma_{ij} + \lambda_{ik} + \epsilon_{ijk} \end{aligned}$$

where Y_{ijks} is the response from the sth test at the jth run from the ith batch at the kth occasion, $\delta_i \sim N(0, \sigma^2_\delta)$, $i = 1, \ldots, 20$, is a random batch effect, $\gamma_{ij} \sim N(0, \sigma^2_\gamma)$, $j = 1, \ldots, n_i$, is a random effect for the orders within batch i, $\lambda_{ik} \sim N(0, \sigma^2_\lambda)$, $k = 1, \ldots, 5$, is a random effect for the combinations of batches and occasions, $\epsilon_{ijk} \sim N(0, \sigma^2_\epsilon)$ is a random oven run effect, $s = 1, 2, 3$ denotes the test and all random variables are independent.

For Bayesian analysis we redefine the notations of random effects corresponding to model (3.5) as $\delta_i \sim N(0, \tau_\delta)$, $\gamma_{ij} \sim N(0, \tau_\gamma)$, $\lambda_{ik} \sim N(0, \tau_\lambda)$, and $\epsilon_{ijk} \sim N(0, \tau_\epsilon)$, where $\tau_\delta = 1/\sigma^2_\delta$, $\tau_\gamma = 1/\sigma^2_\gamma$, $\tau_\lambda = 1/\sigma^2_\lambda$, and $\tau_\epsilon = 1/\sigma^2_\epsilon$ are the precision pa-

rameters. We can define intra-class correlations as $\rho_\gamma = \frac{\sigma^2_\gamma}{\sigma^2_\gamma+\sigma^2_\epsilon}$, $\rho_\lambda = \frac{\sigma^2_\lambda}{\sigma^2_\lambda+\sigma^2_\epsilon}$ and $\rho_\delta = \frac{\sigma^2_\delta}{\sigma^2_\delta+\sigma^2_\gamma+\sigma^2_\lambda+\sigma^2_\epsilon}$ or equivalently in terms of precision parameters as $\rho_\gamma = \frac{\tau_\epsilon}{\tau_\gamma+\tau_\epsilon}$, $\rho_\lambda = \frac{\tau_\epsilon}{\tau_\lambda+\tau_\epsilon}$, and $\rho_\delta = \frac{\tau_\gamma\tau_\epsilon}{\tau_\gamma\tau_\epsilon+\tau_\gamma\tau_\delta}$ respectively. We use the same prior distributions for the intercepts and the fixed effects as before. The prior for the batch variance component σ^2_δ is assumed to be $\sigma_\delta \sim$ Half-Cauchy (where small number of units involved, for instance, as number of batches is moderate, Half-Cauchy works better [Gelman, 2006]) and other variance components are taken into account through intra-class correlations, for instance, $\rho_t \sim$ beta(2.5, 2.5), $t = \gamma, \quad \lambda, \quad \delta$.

We used four different contrasts to capture the effects of categorical factor 'Coating'. The first contrast was 'Solvent-based coating 1 vs 2' (C_1), the second one was 'Water vs solvent-based', i.e. coating 3 vs coating 1 and 2 (C_2), the third one was 'UV coat vs traditional', i.e. coating 5 vs coatings 1, 2, and 3 (C_3), and the last one was 'Low end coat vs rest' which compares the low-end coating 4 with other high-end coatings (C_4).

The final model was obtained by using a manual stepwise forward selection starting from an initial model involving the main effects of 12 factors. Gradually, we increased the complexity by including interaction terms in the model and compared the models with DIC values.

We included interactions of the additives to the polypropylene and the gas plasma treatment factors with coatings to quantify the extent to which the main effects differ across different types of coating. The results from final model appears in Table 3.13.

The combined analysis confirms many of the conclusions drawn from the separate analyses of the coatings. The outputs were more or less similar to those obtained in classical analyses. Most main effects and contrasts appeared to be important. The interaction of UV with other terms was ignorable in the Bayesian analysis though it was significant in the classical analysis. However, the main effect lubricant and many interaction terms, e.g. Ethylene×Power, C2×Ethylene, C2×EPDM, C2×Time and C4×Time were newly detected by the Bayesian analysis.

Table 3.13: Classical and Bayesian estimates obtained from combined analysis of all coatings using a mixed cumulative logit model.

	Likelihood-based Method			**Bayesian Method**		
Effect	Estimate	SE	P-value	Mean	SD	95% CI
Intercept 1	4.502			5.720		
Intercept 2	3.682			4.731		
Intercept 3	2.450			3.250		
Intercept 4	0.976			1.488		
Intercept 5	-1.039			-0.939		
C2	1.218			2.330	0.460	(1.438 , 3.262)
C3	-1.707			-2.843	0.419	(-3.696 , -2.033)
C4	-4.976			-7.059	0.558	(-8.185 , -5.997)
EPDM	0.982	0.221	0.001	1.317	0.373	(0.591 , 2.038)
Ethylene	1.130	0.220	0.000	1.805	0.371	(1.057 , 2.556)
Talcum	1.252	0.268	0.000	1.877	0.433	(1.056 , 2.739)
Mica	0.887	0.274	0.007	1.322	0.459	(0.412 , 2.252)
UV	0.056	0.217	0.801			
Lubrica				0.983	0.357	(0.299 , 1.712)
Power	0.139	0.189	0.466	0.077	0.257	(-0.436 , 0.582)
Time	1.889	0.195	< 0.000	2.765	0.291	(2.198 , 3.351)
Type of Gas	2.066	0.255	<0.000	2.823	0.357	(2.136 , 3.532)
Act. gas	-0.705	0.189	0.000	-0.933	0.257	(-1.451 , -0.432)
EPDM×Talcum				-0.455	0.368	(-1.195 , 0.273)
EPDM×Lubricant				-0.376	0.370	(-1.117 , 0.360)
Power×Type of Gas	0.824	0.291	0.006	1.203	0.395	(0.444 , 2.004)
Power×Act. gas				0.326	0.296	(-0.261 , 0.906)
Power×Time				0.497	0.305	(-0.078 , 1.114)
Time^2	-0.843	0.411	0.044	-1.243	0.566	(-2.342 , -0.110)
EPDM×Time				-0.613	0.259	(-1.132 , -0.120)
EPDM×Type of gas				-0.667	0.333	(-1.324 , -0.008)
Ethylene×Power				-0.538	0.259	(-1.060 , -0.037)
Ethylene×Type of Gas				-0.431	0.338	(-1.103 , 0.219)

Continued on Next Page...

Table 3.13 – Continued

	Likelihood-based Method			**Bayesian Method**		
Effect	Estimate	SE	P-value	Mean	SD	95% CI
C2×EPDM				1.018	0.429	(0.176 , 1.879)
C2×Ethylene				1.095	0.455	(0.224 , 2.011)
C2×Time				1.167	0.424	(0.367 , 2.051)
C3×UV	-0.698	0.244	0.006			
C3×Ethylene	-0.408	0.245	0.104	-0.764	0.402	(-1.556 , 0.031)
C3×Power	0.646	0.259	0.013	1.120	0.363	(0.413 , 1.859)
C3×Time	-0.563	0.261	0.032	-1.007	0.375	(-1.734 , -0.255)
C3×Type of gas	-0.736	0.345	0.034	-1.194	0.471	(-2.132 , -0.291)
C3×Act. gas	0.672	0.259	0.010	0.994	0.353	(0.306 , 1.702)
C4×EPDM				-0.673	0.472	(-1.635 , 0.233)
C4×Ethylene	-0.573	0.294	0.0547	-0.940	0.471	(-1.887 , -0.053)
C4×Time				-1.064	0.460	(-1.981 , -0.155)
C3×EPDM×Ethylene	-0.614	0.245	0.016	-0.966	0.406	(-1.797 , -0.184)
Batch (σ^2_δ)	0.345	0.372		1.180	0.955	(0.160 , 3.717)
Order (σ^2_γ)	0.976	0.438		1.530	0.706	(0.433 , 3.155)
Batch×Occasion(σ^2_λ)	0.220	0.340		1.503	0.731	(0.394 , 3.176)
Run (σ^2_ϵ)	4.435	0.591		8.497	1.297	(6.192 , 11.300)

The variance component estimates for the final model are shown at the bottom of the Table 3.13. The Bayesian estimates differ substantially from the classical estimates. Perhaps reasons are hidden in the SAS estimation procedures and we will discuss this briefly in Section 3.8. Yet, both methods indicate that there is some batch to batch variation and the estimate for σ^2_λ implies that the batch to batch variation was slightly different between occasions. By far the highest variance component estimate (σ^2_ϵ) is due to runs, which provides strong evidence that the three repeated observations are strongly correlated. This was also revealed by the separate analyses of coatings.

In the combined analysis, there is no hidden assumption that there is no inter-main plot variance as all of the variance components could be estimated in the classical methods. Yet, the added value of the Bayesian analysis is that it circumvents some of the technical problems of the REML-GLS method. In classical methods, it is necessary to decide which optimization method to use, whether or not to use the Kenward-Roger method for estimating the standard errors of the fixed effects and which method to use for calculating degrees of freedom for the hypothesis tests [Gilmour and Goos, 2009]. Another advantage of the Bayesian analysis could be if we are unsure about the parameter estimates, at least Bayesian analysis would be a basis to compare the estimates with their classical counterparts and thereby enable the researchers to identify the important factors for the polypropylene experiment.

The main advantage of the combined analysis is that it helps the experimenters to see whether the factor effects differ from coating to coating. If the interaction effects between factors and coatings would all have been insignificant, the information in the data could have been pooled across the different coatings to acquire more precise conclusions about the remaining factors. Some of the interactions of factors and coatings appeared to be important, e.g. C2×EPDM, C3×Power, C3×Time, C4×Ethylene, in the combined analysis. These implied that the effect of EPDM differs between water-based and solvent-base coatings; Power and Time vary from 'UV coat vs traditional' and Ethylene varies with 'Low end coat vs rest'.

3.8 Investigation of Variance Components with Different Priors

The choice of non-informative prior distribution for variance parameters can a have a great effect on inferences, especially when the number of groups is small or the group-level variance, e.g. σ_δ^2, is close to zero. The inverse-gamma(θ, θ) model does not have any proper limiting posterior distribution. As a result, posterior inferences are sensitive to θ – it cannot simply be comfortably set to a low value such as 0.001 [Gelman, 2006].

The uniform(0, B) model yields a proper limiting density as B $\longrightarrow \infty$ as long as the number of groups is at least 3. Thus, for a finite but sufficiently large B, inferences are not sensitive to the choice of B. For our problems, the number of batches or the other groups, e.g. due to runs, orders, Batch×Occasion, are at least 20. Therefore, we tried uniform priors for hierarchical standard deviations e.g. $\log(\sigma_i) \sim \mathrm{U}(-20, 20)$ or simply $\sigma_i \sim \mathrm{U}(0, 25)$.

Half-t prior distributions are more flexible and have better behaviour near 0 (i.e if variance components are near to 0) compared to inverse-gamma family [Gelman, 2006]. As in classical methods some of the variance components were estimated to be zero (e.g. for coating 4, coating 5 in mixed cumulative logit analyses) we assume the half-Cauchy prior, which is special case of half-t family of prior distributions, for the standard deviation parameter. For coating 2 the classical estimates of σ_δ^2 and σ_ϵ^2 are estimated to be 1.2507 and 3.7213 respectively and their Bayesian counterparts obtained by assuming different prior distributions for variance components are presented in Table 3.14.

All the sets of variance component estimates are more or less similar except the first one shown in Table 3.14. The first one differs from others due to assumed prior for batches. The estimates of variance components are quite stable for models 4-8. It is difficult to say which set of Bayesian estimates are precise than others. Even when we used all non-informative priors for both of the variance components $\rho \sim$ beta(1, 1) and log $\sigma \sim$ U(-20, 20), we found almost similar estimated variance components. However, our assumption is that classical estimates are deviated from

Table 3.14: Investigation of variance components with different priors

Model	Priors			Estimated Variance Components	
	ρ	Batch	Run	σ^2_δ	σ^2_ϵ
1	-	$\tau_2 \sim$ gamma$(5, 5)$	$\tau_1 \sim$ gamma$(1, 1)$	1.512	5.477
2	beta(5.0, 5.0)	-	$\tau_1 \sim$ gamma$(1, 1)$	4.131	5.050
3	beta(2.5, 2.5)	-	$\tau_1 \sim$ gamma$(1, 1)$	3.805	5.141
4	beta(2.5, 2.5)	-	$\log(\sigma) \sim$ U$(-20, 20)$	3.933	5.670
5	beta(2.5, 2.5)	-	$\log(\sigma) \sim$ U$(-5, 5)$	4.043	5.677
6	beta(2.5, 2.5)	-	$\sigma \sim$ U$(0, 25)$	4.070	5.964
7	beta(2.5, 2.5)	-	$\sigma \sim$ Half-Cauchy	4.025	5.919
8	beta(1.0, 1.0)	-	$\log \sigma \sim$ U$(-20, 20)$	3.482	5.752
9	beta(1.0, 4.0)	-	$\log \sigma \sim$ U$(-20, 20)$	2.398	6.295

the true situations. Perhaps some reasons are explored by Zhang et al. [2011]. They evaluated performances of different methods/procedures in SAS and R packages. Unlike special cases of mixed models, computation of MLE is difficult and approximate methods have been proposed and implemented in various packages including SAS. That is why we are unsure whether the estimates, particularly variance components estimates, computed by SAS or by other packages are precise. We found the estimates in Bayesian methods are similar assuming weakly informative or even non-informative priors for variance components. Therefore, from our intuition and evidence we conclude that estimate of batch-to-batch variation is not less than 3 and the estimate of variation due to runs is around 5 or more for coating 2. Pettit [1986] indicated how much effects of other priors have can be assessed comparing with a reference prior. If we consider priors of variance components in model 8 as reference priors, since both are noninformative, we can see how much deviations of other variance components have.

3.9 Convergence Diagnostics

We used the convergence diagnostics for the parameters of all the selected models for the five coatings, however present results, as an example, only for coating 2.

To monitor convergence, we begin with the Monte Carlo error (MC error) for some of the parameters from the best cumulative logit model for coating 2. A rule of thumb regarding convergence is that MC errors should be less than 5% of the cor-

responding posterior standard deviations (SD). In coating 2 analysis, all of MC errors were less than 5% of the respective SDs (see Table 3.15).

Table 3.15: Standard deviation and Monte Carlo error (MC Error) in Bayesian analysis of coating 2

Effect	Mean	SD	MC Error
Intercept 1	7.5460	0.9648	0.0204
Intercept 2	5.9200	0.8565	0.0175
Intercept 3	3.7400	0.7548	0.0165
Intercept 4	1.6980	0.7044	0.0158
Intercept 5	-0.4783	0.6893	0.0142
EPDM	0.8028	0.5776	0.0085
Ethylene	2.0630	0.5898	0.0087
Talcum	2.0830	0.6970	0.0145
Mica	1.1990	0.7062	0.0152
Power	-0.3401	0.3520	0.0046
Time	2.5850	0.4100	0.0058
Gas Type	3.3210	0.5668	0.0089
Act. gas	-0.7751	0.3615	0.0051
EPDM×Ethylene	-0.7023	0.5633	0.0092
EPDM×Talcum	-1.0490	0.5833	0.0088
Power×Gas Type	1.8530	0.5824	0.0088
Power×Act. gas	0.9698	0.4010	0.0051
Power×Time	1.3340	0.4420	0.0065
Ethylene×Power	-0.6436	0.3507	0.0051
Mica×Act.gas	0.7440	0.3543	0.0052
Batch (σ^2_δ)	4.1310	2.2160	0.0328
Run (σ^2_ϵ)	5.0500	1.7070	0.0278

We find in Figure 3.3 that kernel densities of posteriors corresponding to the factors EPDM, ethylene, batch and run are approximately smooth and unimodal. There is no visible non-convergence problem in the kernel densities.

The plots in Figure 3.4 are examples of multiple chains for which convergence looks reasonable for all the factors. However, for a parameter of interest if the level of autocorrelation is strong then a traceplot will be poor diagnostic for for convergence.

There were some evidence of autocorrelation in the data before thinning, however, autocorrelation effects disappeared when we thinned the data as 20th and these

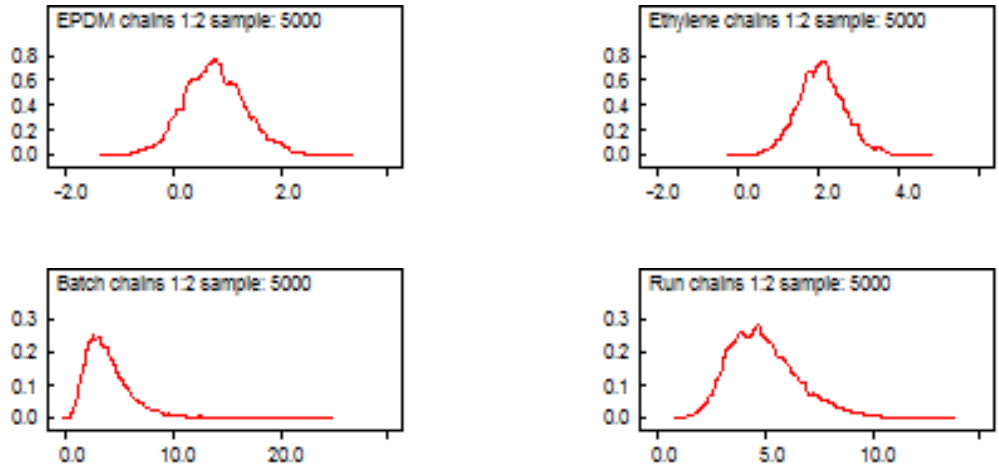

Figure 3.3: Kernel density of few a parameters of the best model for coating 2; EPDM (top left), Ethylene (top right), Batch (bottom left), Run (bottom right).

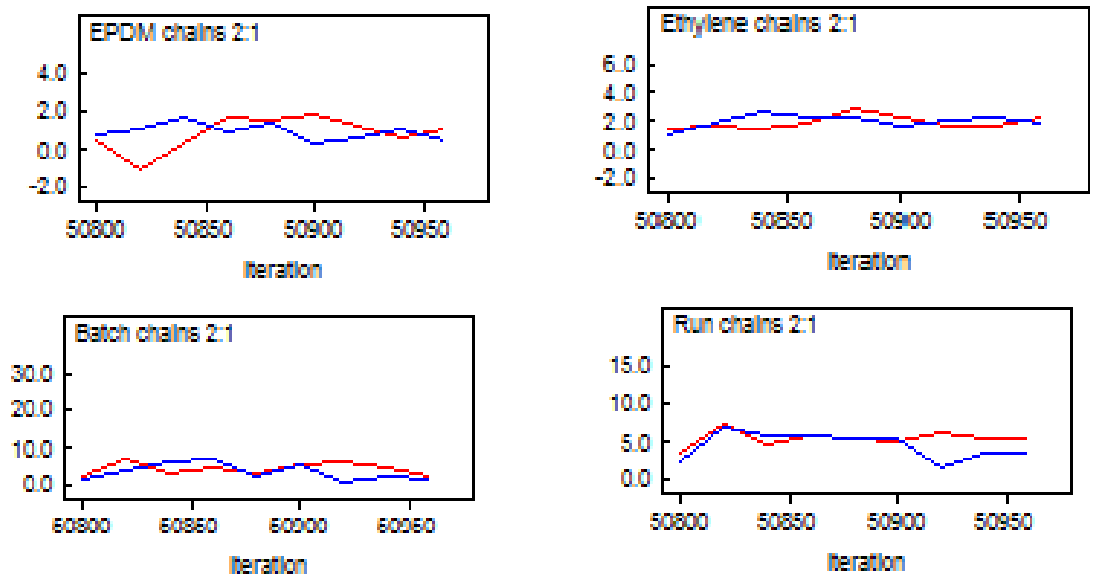

Figure 3.4: Trace plots of a few parameters of the best model for coating 2; EPDM (top left), Ethylene (top right), Batch (bottom left), Run (bottom right).

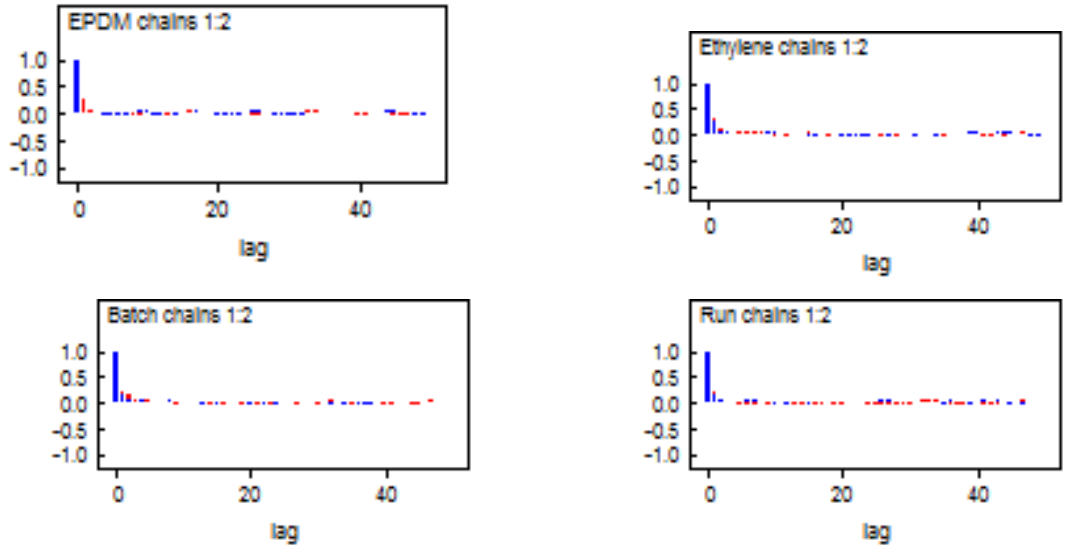

Figure 3.5: Autocorrelation status of few parameters after thinned as 20th for coating 2; EPDM (top left), Ethylene (top right), Batch (bottom left), Run (bottom right).

are reflected in Figure 3.5. In Figure 3.6 we also find the well mixing of chains which provides evidence of convergence.

The Figure 3.7 displays the examples of Gelman-Rubin statistics (R). We started with two sets of initial values for all parameters. It seems that the R statistics are approximately equal to 1 for all the parameters. Therefore, we may conclude that convergence is achieved for all parameters.

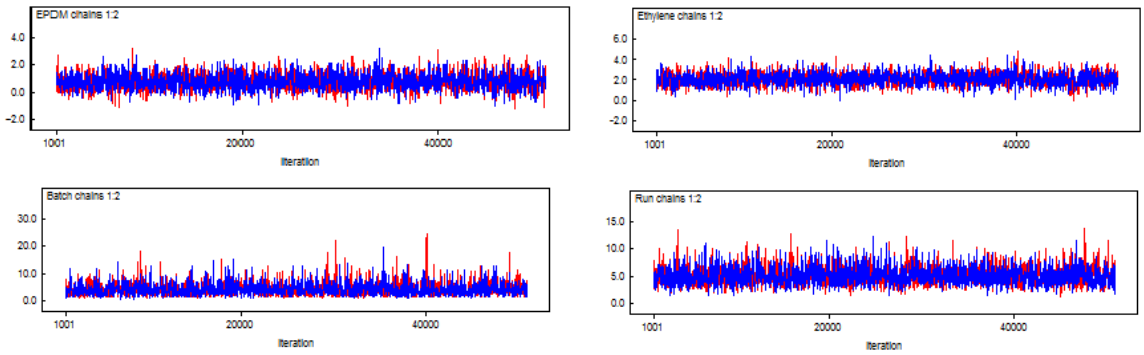

Figure 3.6: History plot of few parameters of the best model corresponding to coating 2; EPDM (top left), Ethylene (top right), Batch (bottom left), Run (bottom right).

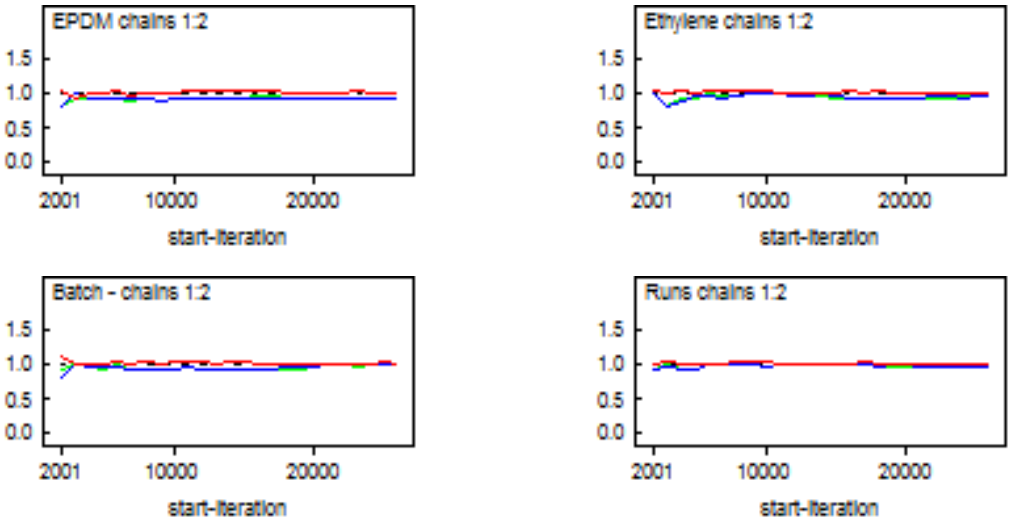

Figure 3.7: Gelman-Rubin statistic of few parameters of the best model corresponding to coating 2; EPDM (top left), Ethylene (top right), Batch (bottom left), Run (bottom right).

In our study we find that the parameters of the selected models passed the convergence criteria for all the coatings in cumulative logit analysis.

3.10 Profile Likelihood, Confidence Intervals and Simulation Studies

Likelihood based confidence intervals (CIs) have been computed from Wald statistic, profile likelihood and bootstrap based methods in the binary logit analysis of polypropylene experiment. These confidence intervals have been compared with the Bayesian credible intervals which are shown in Table 3.16. The model that is concerned with the computation of likelihood based CIs and Bayesian credible intervals is similar to model (3.3) which has five fixed effects EPDM, ethylene, talcum, time and gas type and two random effects batch and run under coating 2 data. To reduce complexity and to save computational time we did not consider any interaction terms in the current model. The priors assumed during computation of Bayesian credible intervals are $\beta_i \sim N(0.001, 0.1)$, $i = 1, 2, \ldots, 5$, $\tau_\epsilon \sim \text{Gamma}(5, 5)$ and $\rho \sim \text{U}(0.001, 0.9)$, where $\tau_\epsilon = 1/\sigma_\epsilon^2$, $\tau_\delta = 1/\sigma_\delta^2$ and $\rho = \frac{\tau_\epsilon}{\tau_\epsilon + \tau_\delta}$. It is found in Table 3.16 that the fixed effect estimates in likelihood and Bayesian methods are

similar. Interestingly, all likelihood based point estimates except variance component due to runs (σ^2_ϵ) are within Bayesian credible intervals and all Bayesian point estimates are within Wald, profile likelihood and bootstrap based confidence intervals (CIs). However, Bayesian credible intervals have less width than likelihood based CIs.

This scenario of Bayesian credible intervals in polypropylene experiment is slightly different than the credible intervals shown in Table 2.18 and 2.19 of Chapter 2 where Bayesian credible intervals were slightly wider than likelihood based CIs. In polypropylene experiment, Bayesian intervals have less width than that of likelihood based CIs, because perhaps comparatively polypropylene experiment has larger data and outcomes are therefore influenced by data as non-informative or weakly informative priors have been used. On the other hand in fuel economy experiment data could not play the leading role alone as the experiment had smaller data. Further, lower limits of profile likelihood and bootstrap based intervals in Table 3.16 show that variance component due to batch (σ^2_δ) could be zero, whereas Bayesian method enables non-zero estimates of (σ^2_δ). Therefore, implementation of Bayesian methods might ensure non-zero estimates of variance components and it is evident that Bayesian point estimates are performing at least as well as likelihood based estimates.

Table 3.16: Likelihood and Bayesian estimates with 95% intervals under different methods

Parameter	Likelihood Method				Bayesian Method	
		95% CI				95% CI
	Estimate	Wald	Profile	Bootstrap	Estimate	
α	5.45	(2.43, 8.46)	(3.21, 9.81)	(3.91, 12.72)	3.69	(2.28, 5.38)
β_1	1.60	(-0.07, 3.27)	(0.09, 4.14)	(0.13, 5.84)	1.10	(0.00, 2.25)
β_2	2.53	(0.66, 4.40)	(1.02, 5.14)	(1.14, 7.54)	1.75	(0.51, 3.18)
β_3	2.62	(0.50, 4.75)	(0.91, 5.69)	(1.01, 7.25)	1.69	(0.45, 3.19)
β_4	3.36	(1.25, 5.47)	(1.74, 6.32)	(1.95, 7.80)	2.36	(1.39, 3.58)
β_5	4.54	(1.43, 7.65)	(2.18, 9.14)	(2.60, 11.43)	3.32	(1.89, 5.17)
σ^2_δ	2.64	-	(0.00, 19.89)	(0.00, 15.13)	3.63	(0.28, 11.34)
σ^2_ϵ	16.09	-	(4.97, 62.88)	(7.08, 128.56)	5.39	(1.75, 12.38)

Simulation studies under mixed binary logit models have been performed at a limited scale as simulation with a mixed model and bigger data is very time consuming. For instance, simulation studies with profile likelihood and bootstrap based

Table 3.17: Simulated performance of maximum likelihood and Bayesian estimates under binary logit model in polypropylene experiments assuming true parameter values as $\beta_0 = 3.6$, $\beta_1 = 1$, $\beta_2 = 1.7$,$\beta_3 = 1.5$,$\beta_4 = 2.3$,$\beta_5 = 3.2$, $\sigma^2_\delta = 3$, $\sigma^2_\epsilon = 4.2$; and sample size n=300

Parameter	Mean/Posterior Mean			Median/Posterior Median			Coverage Probability	Average width of 95% CI	Median width of 95% CI
	Bias	% Relative Bias	RMSE	Bias	% Relative Bias	RMdSE			
MLE and Wald based intervals									
α	112.193	3116.48	3532.878	0.059	1.64	0.639	0.888	1061.405	3.252
β_1	0.091	9.13	0.723	-0.002	-0.19	0.425	0.869	2.195	2.087
β_2	0.127	7.47	0.750	0.021	1.24	0.459	0.895	2.354	2.240
β_3	0.128	8.53	0.798	0.017	1.14	0.464	0.875	2.489	2.373
β_4	0.178	7.73	0.762	0.077	3.35	0.434	0.897	2.400	2.274
β_5	224.111	7003.47	7065.757	0.069	2.14	0.625	0.897	2119.498	3.431
σ^2_δ	-0.697	-23.22	3.079	-1.493	-49.76	1.926	-	-	-
σ^2_ϵ	1.158	27.57	4.644	-0.070	-1.66	1.980	-	-	-
MLE and profile likelihood based intervals									
α	726.106	20169.62	15684.021	0.105	2.92	0.655	0.916	∞	3.651
β_1	0.107	10.66	0.729	0.028	2.84	0.412	0.925	2.670	2.452
β_2	0.105	6.19	0.783	0.016	0.97	0.444	0.914	2.805	2.548
β_3	0.121	8.05	0.787	0.040	2.69	0.472	0.922	2.976	2.709
β_4	0.166	7.22	0.758	0.041	1.77	0.422	0.922	2.699	2.413
β_5	1451.875	45371.11	31368.152	0.131	4.09	0.595	0.937	∞	3.740
σ^2_δ	-0.305	-10.18	2.691	-0.935	-31.15	1.544	0.528	∞	2.698
σ^2_ϵ	0.998	23.77	5.014	-0.383	-9.11	1.954	0.381	∞	2.714
MLE and bootstrap based intervals									
α	0.452	12.56	1.411	0.565	15.70	0.891	1	6.506	6.168
β_1	0.522	52.22	0.945	0.209	20.85	0.304	1	3.364	3.285
β_2	0.312	18.34	0.738	0.485	28.53	0.702	1	3.724	3.693
β_3	0.466	31.06	1.020	0.619	41.27	1.005	1	4.036	4.107
β_4	0.246	10.72	0.851	0.245	10.64	0.723	1	3.353	3.537
β_5	0.114	3.58	0.933	-0.047	-1.46	0.556	1	7.047	4.839
σ^2_δ	0.365	12.18	3.804	-1.057	-35.23	2.191	1	3.007	3.347
σ^2_ϵ	1.550	36.90	3.445	1.286	30.61	1.790	1	4.396	4.383
Priors: $\beta_0 \sim N(0.001, 0.1)$, $\beta_i \sim N(0.001, 0.1), i = 1, 2, \ldots, 5$, $\tau_\epsilon \sim Gamma(5, 5)$, $\rho \sim U(0, 0.9)$									
α	-0.953	-26.48	1.084	-0.969	-26.92	0.969	0.725	2.493	2.470
β_1	-0.179	-17.93	0.518	-0.181	-18.10	0.342	0.932	2.130	2.113
β_2	-0.318	-18.68	0.545	-0.316	-18.58	0.382	0.927	2.156	2.147
β_3	-0.281	-18.77	0.595	-0.289	-19.28	0.435	0.933	2.283	2.267
β_4	-1.275	-55.45	1.367	-1.297	-56.39	1.297	1.000	8.459	8.416
β_5	-2.199	-68.71	2.250	-2.176	-67.99	2.176	0.999	8.458	8.363
σ^2_δ	-0.195	-4.15	1.615	-0.473	-10.07	1.182	0.964	8.022	7.603
σ^2_ϵ	-0.118	-3.94	1.660	-0.447	-14.89	1.212	0.964	8.566	8.131

methods took several weeks to accomplish only 100 simulations. As before, a mixed model with five fixed effects corresponding to factors namely EPDM, ethylene, talcum, time and gas type and two random effects batch and run has been considered for simulation studies. In simulation studies true parameters were $\beta_0 = 3.6$, $\beta_1 = 1$, $\beta_2 = 1.7$, $\beta_3 = 1.5$, $\beta_4 = 2.3$, $\beta_5 = 3.2$, $\sigma^2_\delta = 3$, and $\sigma^2_\epsilon = 4.2$ which were conjectured from individual studies of likelihood and Bayesian methods for coating 2 with five fixed and two random effects.

Table 3.17 shows that all likelihood estimates are upwardly biased whereas Bayesian estimates have downward bias. Thus likelihood methods over estimate the parameters and Bayesian method under estimates the parameters. In case of Wald based computations, we do not have coverage probabilities for σ^2_δ and σ^2_ϵ as no confidence intervals are available for the variance components in Wald based method because sampling distribution of variance components will be a normal distribution whereas the assumption of Wald based computation is the asymptotic normality of parameter estimates. Mean and median based estimates have similar pattern in the results of bias and relative bias. Wald and profile likelihood based methods have coverage lower than 95% nominal coverage probabilities. However, bootstrap based method shows over coverage as all equal to 1 which is not acceptable as over coverage consequences type II errors more. For Bayesian cases, coverage probabilities are acceptable except α and β_5 which have under and over coverages respectively. For some of the parameters, for example α, σ^2_δ and σ^2_ϵ average widths of 95% CIs are ∞ in the profile likelihood based method. This happens as in some simulations upper limit of the CI were ∞. However, it seems that simulation results under bootstrap based and Bayesian methods are close though bootstrap based estimates are unreliable because of having over coverage. Yet, considering all aspects we may end up with the notion that Bayesian techniques with appropriate choice of priors might enable comparatively better estimates in polypropylene experiment particularly for variance components.

3.11 Conclusion

In this study we have investigated Bayesian models for the data from the polypropylene experiment and compared Bayesian and classical estimates obtained from split-

plot and multi-stratum industrial experiments. We tried mixed binary logit models when responses were dichotomous and mixed cumulative logit models when responses were naturally ordered categorical. For all coatings, we have used deviance information criterion (DIC) to select the best models in Bayesian methods.

In binary logit analysis of polypropylene experiment, the factors time and gas type appeared to be important for all coatings. EPDM and ethylene found to be important for all coatings except coating 4. There were convergence problems while we were including certain interaction terms in the model both in classical and Bayesian approaches. Undoubtedly a large part was due to the nature of categorical more specifically binary responses. The variance component due to batch was estimated as zero for coating 3 and coating 4 in classical method. However, Bayesian method assuming some weakly informative priors enables non-zero estimation of variance component due to batch. In cumulative logit analysis of ordered responses time and gas type were important for all coatings both in classical and Bayesian methods and a few of interactions e.g. power×act.gas and power×gas type were found non-negligible in some coatings. Further, the variance component due to batch was estimated as zero in cumulative logit analysis for coatings 1, 4 and 5. However, zero variance component is not unusual for likelihood based methods as shown in a brief analysis under profile likelihood and bootstrap based methods (see Table 3.16). On the other hand, Bayesian methods provide non-zero estimates of variance components due to batch for cumulative logit analysis of all coatings.

In our analysis we followed the stepwise manual forward selection procedures. Chipman and Hamada [1996] used manual backward elimination method to select the variables. We could not start backward elimination for two reasons. Firstly, it was not feasible for binary responses due to having convergence problems because of the existence of certain interactions in the model. Secondly, we did not attempt for ordered categorical response data analysis due to consuming longer time to terminate or even sometimes impossible for some coatings (e.g. coating 4) probably for the same reasons as before in binary data analysis. Our model selection approach was rather informal and other approaches could be automated Bayesian variable selection [George and McCulloch, 1993] or predictive approach [Box, 1980]. Further, recently Tan and Wu [2013] proposed a global and local search (GLS) algorithm to find models with high posterior probabilities and to estimate posterior model

probabilities in fractionated split plot experiments.

We have compared interval estimates of parameters under likelihood and Bayesian methods for coating 2 shown in Table 3.16. In likelihood based approach Wald statistic, profile likelihood and bootstrap based methods have been considered to construct confidence intervals. Bayesian credible intervals appeared to perform better than likelihood based CIs in terms of less wide intervals and non-zero lower limit of the variance components. Simulation under binary logit analysis have been performed at a limited scale due to time constraints. Also, the nature of binary data and non-informative priors does not allow experimenters much to do trial and errors in simulation studies under binary logit analysis. Simulation studies under cumulative logit models will be investigated later in future studies.

One limitation of the current study is that we did not test the proportionality assumptions in Bayesian approach, as the methods are yet to be developed. However, one way of getting round this issue could be fitting separate logistic models across the categories of an outcome variable assuming priors for coefficients corresponding to different levels of outcome categories. The comparison of the separate logistic fits for a model can provide supportive information regarding the plausibility of parallelism for the data. However, it was also difficult in classical method to detect departures from proportional odds which could not be fixed by changing some other aspect of the model.

In this chapter, we have shown Bayesian methods implemented by MCMC techniques through WinBUGS for binary and ordered categorical responses. Our intention was not to establish Bayesian methods as superior to classical methods, rather to explore the gap left by either classical or Bayesian methods during analysis of binary and cumulative logit models in polypropylene data. To sum up the discussion, perhaps Bayesian method give better results for analyzing polypropylene data, particularly in case of variance component estimation. This Bayesian techniques implemented here could be very useful for industrial and biomedical researches where binary and ordered categorical responses arise naturally.

Chapter 4

Optimal Design for Categorical Data Minimizing the Probability of Separation

4.1 Optimal Design in Statistics

Optimal design of experiments is an important subfield in statistics. Optimal design ideas are widely used in many disciplines and continually applications are increasing in new fields. One of the reasons behind this driving force is the ever-increasing cost of running experiments or field projects [Berger and Wong, 2005].

In design of experiments, an optimal design is an experimental design that is optimal with respect to a statistical criterion. Optimal design is also called an optimum design in literature. For estimating statistical models, optimal designs allow parameters to be estimated with minimum variance. A non-optimal design requires a greater number of experimental runs to estimate the parameters with the same precision as an optimal design. Thus, an optimal design can reduce the costs of experimentation.

Statistical experimental designs are judged using optimality criteria. The most im-

portant design criterion in applications is that of D-optimality criterion, in which generalized variance (where generalized variance is the inverse of the determinant of information matrix) of the parameter estimates is minimized [Atkinson et al., 2007]. The generalized variance or variance covariance matrix of parameters determines the shape and form of the confidence ellipsoid of the parameters. Again the volume of the confidence ellipsoid is inversely proportional to the square root of the information matrix. Thus, the shape as well as the volume of the confidence region depends on the information matrix. Designs which maximize determinant of information matrix or minimize the determinant of covariance matrix of parameters are called D-optimal. The popular optimality criteria including D-optimality are discussed extensively in Atkinson et al. [2007]. In this Chapter D-optimality criterion will be compared with the newly developed probability criteria namely P_s- and DP_s-optimality which have been proposed to minimize a statistical computational annoyance known as separation.

4.2 Separation Problem in Categorical Data Analysis and Non-existence of Maximum Likelihood Estimates

Categorical data consists of variables with a finite number of discrete values. They arise in a number of ways, for instance, by simple counts; binary variables-yes, no; unordered multinomial- christian, jew, muslim, atheist; ordered variable-Likert's five-point scale.

Categorical data can be analyzed by numerous statistical methods. During analysis of categorical data often parameter estimates are not available due to a problem known as separation which is considered as a somewhat incomprehensible computational annoyance. In separation, a response variable is separated into two categories at a point of a covariate or a linear combination of covariates. For example, Y is a binary response variable and X is an independent variable, then if Y is separated at X=0, more specifically Y=0 for X$\leq$0 and Y=1 for X>0 then this is considered as a separation problem. The separation problem causes the non-existence of maximum likelihood estimates (MLEs) of parameters in logistic

regression. The same is true for related models including probit and many other similar models for binary and multinomial regression. In some context, separation is often quite good at classifying observations, but inferences about parameters are not reliable always. For example, in biostatistics, if we can determine exactly at which level of a risk factor (X) the patient become diseased (Y=1) or not diseased (Y=0), then it is useful in medical research though Y could be separated at that level of X. Separation occurs primarily in small samples. With increasing sample size, the probability of observing a set of separated data points tends to zero, no matter what the sampling design.

4.2.1 Types of Separation

Albert and Anderson [1984] classify logistic regression data sets into three mutually exclusive and exhaustive categories: complete separation, quasi-complete separation, and overlap. The maximum likelihood estimates exists only for overlapped data. Let us define a response variable Y, a vector of covariates $\mathbf{X}$ and corresponding vector of coefficients $\boldsymbol{\alpha}$. Then we can define two frequently used terms as follows

Complete separation occurs whenever there exists some vector of coefficients $\boldsymbol{\alpha}$ such that $Y_i = 1$ if $\boldsymbol{\alpha}'\boldsymbol{X_i} > \mathbf{0}$ and $Y_i = 0$ if $\boldsymbol{\alpha}'\boldsymbol{X_i} \leq \mathbf{0}$. In other words, complete separation occurs whenever a linear function of $\mathbf{X}$ can generate perfect predictions of Y

Quasi-complete separation occurs when there exists some coefficient vector $\boldsymbol{\alpha}'$ such that $Y_i = 1$ if $\boldsymbol{\alpha}'\boldsymbol{X_i} \geq \mathbf{0}$ and $Y_i = 0$ if $\boldsymbol{\alpha}'\boldsymbol{X_i} \leq \mathbf{0}$, and equality holds for at least one case in each category of the dependent variable.

Data which are neither completely or quasi-completely separated are called overlapped. Given a non-singular $\boldsymbol{X}'\boldsymbol{X}$ matrix a certain degree of overlap is a necessary and sufficient condition for the existence of maximum likelihood estimates in logistic link function for the binomial response. [Silvapulle, 1981].

Table 4.1: Hypothetical examples of separation problem

(a) Data exhibiting complete separation

Y	X
0	-6
0	-5
0	-4
0	-3
0	-2
0	-1
1	1
1	2
1	3
1	4
1	5
1	6

(b) Data exhibiting quasi-complete separation

Y	X
0	-6
0	-5
0	-4
0	-3
0	-2
0	-1
0	0
1	0
1	1
1	2
1	3
1	4
1	5
1	6

4.2.2 Hypothetical Example of Separation

We consider two hypothetical examples of separation in Table 4.1. The complete separation is shown in Table 4.1a and Table 4.1b displays a data set regarding quasi-complete separation. What distinguishes the data set in Table 4.1b is that there are two additional observations, each with x values of 0 but having different values of y.

4.2.3 Separation Problem in the Current Study

We have faced convergence problems during binary data analysis of the polypropylene experiment. One of the reasons could be separation in the data. Allison [2008] gives several small data-sets to explain how complete and quasi-complete separation lead to the non-existence of the maximum likelihood estimator. The separation can be detected by inspecting contingency tables formed by a discrete variable and the response variable. Whenever a cell in such a contingency table contains a zero, the maximum likelihood estimator of the corresponding parameter does not exist.

In the polypropylene experiment, adding the two-factor interaction effect of EPDM and type of gas caused the convergence to fail when analyzing the success of coat-

Table 4.2: Existence of separation problem in the current study

EPDM	Type of gas	Success of coating 0	1	Total
0%	Etching	27	21	48
0%	Activation 1	13	38	51
0%	Activation 2	6	36	42
10%	Etching	0	42	42
10%	Activation 1	14	43	57
10%	Activation 2	22	38	60
	Total	82	218	300

ing 2. The contingency table displayed in Table 4.2 does indeed have a zero cell and is evidence of separation in the polypropylene experiment when trying to fit the interaction EPDM×Type of gas [Goos and Gilmour, 2012]. However, the problem of MLE non-existence relevant to logistic models will be explored further in the next section.

4.2.4 Non-existence of Maximum Likelihood Estimates

The log-likelihood under logistic model is globally concave, meaning that the function has at most one maximum. Unfortunately, there are many situations in which the likelihood function has no maximum, in which cases we say that the maximum likelihood estimate of at least one parameter does not exist. According to Albert and Anderson [1984] nonexistence of the maximum likelihood estimate means absence of a finite maximum. Fienberg and Rinaldo [2007] used the wording "non-existence of the MLE" to signify lack of solutions for the maximum likelihood optimization problem. To illustrate the non-existence of MLE we consider the logit model as $logit(\pi_i) = \beta_0 + \beta_1 X_i$ and take the data sets similar to Allison [2008], presented in Table 4.1. For these data sets, it can be shown that the maximum likelihood estimate of intercept (β_0) is 0. Suppose that X is a dichotomous variable (e.g. X is converted to 0 when $X < 0$ and to 1 when $X \geq 0$) so that the data can be arrayed in a 2×2 table, with observed cell frequencies n_{11}, n_{12}, n_{21}, and n_{22} shown in Table 4.3a. Then maximum likelihood estimate of β_1 is given by the logarithm of the "cross-product" ratio:

$$\hat{\beta}_1 = \ln\left(\frac{n_{11} \times n_{22}}{n_{12} \times n_{21}}\right) \tag{4.1}$$

Table 4.3: Hypothetical examples of overlapped and separated data

(a) Overlapped data

		x	
		0	1
y	0	n_{11}	n_{12}
	1	n_{21}	n_{22}

(b) Data with complete separation

		x	
		0	1
y	0	10	0
	1	0	5

(c) Data with quasi-complete separation

		x	
		0	1
y	0	5	0
	1	15	10

The likelihood estimate is obtainable for overlapped data, for example, shown in Table 4.3a when cell frequencies n_{11}, n_{12}, n_{21}, and n_{22} are non-zero. Fienberg and Rinaldo [2007] demonstrated various aspects of the patterns of zeros that lead to non-existence of MLE and various practical considerations that follow from non-existence. However, contingency table with sampling zeros i.e. with having some cell frequencies as zeros, for instance, in complete separation shown in Table 4.3b and in quasi-complete separation shown in Table 4.3c, MLE does not exist. Because for Table 4.3b by using (4.1) we obtain the slope parameter:

$$\hat{\beta}_1 = \ln\left(\frac{10 \times 5}{0 \times 0}\right) = \infty. \tag{4.2}$$

Again for Table 4.3c by using 4.1 we obtain the slope parameter:

$$\hat{\beta}_1 = \ln\left(\frac{5 \times 10}{15 \times 0}\right) = \infty. \tag{4.3}$$

In both of the cases, the estimate of slope parameter is undefined because the denominator is zero. The same problem would happen if there were a zero in the numerator because logarithm of zero is also undefined. Thus, in general whenever there is a zero cell frequency in a 2×2 table, the MLE of β_1 under logistic regression does not exist. This is by far the most common reason for convergence failure in logistic regression [Allison, 2008]. To illustrate why MLE does not exist, we can also sketch a graph of log-likelihood function against the slope parameter as in Figure 4.1. It is evident from the figure that log-likelihood does not reach a maximum as β_1 increases. It is bounded above by 0. The log-likelihood can be close to 0 by making β_1 sufficiently large. Therefore, there is no maximum likelihood estimate of β_1.

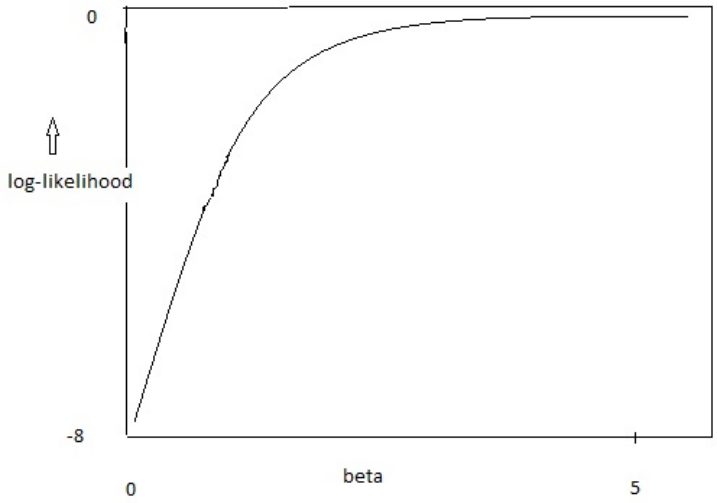

Figure 4.1: Log-likelihood as a function of the slope under separation

Though separation is more likely to occur in small sample, it can occur even in large sample when there are extreme splits on the frequency distribution of either dependent or independent variables. This was also evidenced in simulation studies of binary response model in the polypropylene experiment where a few of complete separations were happened in a large sample.

The non-existence of MLEs has been illustrated under various situations of contingency tables by Haberman [1974] and Fienberg and Rinaldo [2007], but mostly on log-linear models where logistic model is considered under this log-linear family of models.

Haberman [1974] studied extensively the problem of existence, finiteness and uniqueness of maximum likelihood estimates in log linear models, including quantal-response models. He proves a very general theorem (Haberman, 1974, p. 37) on necessary and sufficient conditions for the maximum likelihood estimate to exist. In his terminology, MLE existence means finiteness of the solution. He also demonstrates that for most models, if the maximum likelihood solution exists, it is unique, as a result of the concavity of the likelihood function. Wedderburn [1976] presents a series of sufficient, but not necessary, conditions for the existence, uniqueness and location, on the boundary or not, of maximum likelihood estimates for the parameters of the generalized linear model. Albert and Anderson [1984]

noted that for multinomial logistic regression these conditions are not satisfied in many practically important cases, such as completely and quasi-completely separated data configurations. Though powerful, both Haberman's and Wedderburn's results fall short of providing conditions for nonexistence of maximum likelihood estimates and general practical procedures for identifying infinite parameter values, except in problems with a special structure [Albert and Anderson, 1984]. Thus, although Haberman [1974] gave necessary and sufficient conditions for the existence of MLE, his characterization is non-constructive in the sense that it does not directly lead to implementable numerical procedures and also fail to suggest alternative methods of inference for the case of undefined MLE. Despite these deficiencies, Haberman (1974)'s results have remained all that exist in the published literature [Fienberg and Rinaldo, 2007].

As authors noted that Haberman (1974)'s works are rather theoretical and has not been useful in solving real life problems relevant to non-existence of MLE, it is not surprising that virtually all implemented computational algorithms do not take into account the Haberman's non-constructive characterization of MLE non-existence while dealing with sampling zeros in the contingency tables. For example, in SAS, by default PROC FREQ does not process observations that have zero weights, because these observations do not contribute to the total frequency count, and because any resulting zero-weight row or column causes many of the tests and measures of association to be undefined. The presence of sampling zeros is dealt with by adding small positive quantities to the zero cells to facilitate the convergence of the underlying numerical procedure. However, Fienberg and Rinaldo [2007] illustrate various issues and dangers associated to the usage of very common computational procedures for obtaining the MLE with artificially constructed tables and with a simple real-life example of a non-sparse contingency table (Fienberg and Rinaldo, 2007 p. 3440).

4.3 Existing Ways of Dealing with the Separation Problem

In practice, the researchers handle separation problem separately for the complete and quasi-complete cases.

4.3.1 Solutions for Quasi-complete Separation

In many cases the problem of quasi-complete separation is handled simply by deleting the problem variable whose coefficient did not converge. However, deleting method is not well recommended as deleting variables with strong effects will definitely obscure the effect of those variables, and is likely to bias the coefficients for other variables in the model Allison [2008]. In case of large number of categories for some variables, combining categories might be useful to minimize the cells with zero frequencies and thereby reducing the problem of separation. However, if the dummy or categorical variable represents an irreducible dichotomy, for instance sex, then this solution is not feasible. Allison [2008] demonstrated that a non-informative prior does not work satisfactorily in dealing with separation in Bayesian method. We also found this true having difficulties in convergence during binary analysis of data from polypropylene experiment while we were assuming non-informative priors for fixed effects in the models.

Penalized maximum likelihood estimation (PMLE) proposed by Firth [1993] is widely used to deal with separation problem in small samples. However, Wald tests based on the standard errors for variables causing separation can be highly inaccurate similar to conventional maximum likelihood method. Again, for small sample problems, exact methods can be used for some basic analysis of contingency table rather than using large-sample approximations when their adequacy is in doubt [Agresti, 1992]. As separation is largely a small sample problem, exact logistic regression which was proposed by Cox [1970] may be used for analysis. The coefficient estimates reported with the exact logistic regression method are usually conditional maximum likelihood estimates, even these may not exist when there is separation [Allison, 2008]. Though utility of exact methods diminishes with the increase of sample size, yet we have a plan to make use of exact methods while dealing with separation in future studies.

4.3.2 Solutions for Complete Separation

The complete separation is considerably more difficult to deal with. It is impossible to obtain the likelihood estimates keeping the variable that causes separation in the model. Combining categories might not be useful to solve the problem either. Exact logistic regression might be useful for small samples but not computationally

feasible for large samples [Allison, 2008]. Bayesian estimation may be suitable, but it requires informative prior for the parameter associated with the problem variable, and also results may be sensitive to the choice of that prior distribution. Further as noted earlier, non-informative prior does not work satisfactorily to deal with separation. In practice, the problem of complete separation is handled by deleting the problem variable from the model. That enables one to obtain estimates for the remaining variables, however, the exclusion of the problem variable might lead to biased estimates for the remaining variables.

Considering these pitfalls in the existing methods, we propose optimal design techniques to minimize the probability of separation problem as well as precise parameter estimation in this study.

4.4 Probability of Separation

The maximum likelihood estimate does not exist for the logistic regression data with complete or quasi-complete separation. We want to maximize the probability of the maximum likelihood estimates existence through optimal design techniques, that implies the minimization of the probability of separation. Now we describe the notations to be used for separation probability.

Let Y_i be a response corresponding to ith observation and the associated covariate be X_i , $i = 1, 2, \ldots, n$. The distribution of Y_i is Bernoulli with probability of success π_i. In matrix notation

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

For simplicity in defining the separation probability, let us describe Table 4.4 with Y_i and X_i values arranged in order from the smallest to the largest. In the X_i column there is a cut off point X^* which separates the outcome Y into two groups namely 'success' (S_i) defined by $Y_i = 1|X_i$ and 'failure' (F_i) defined by $Y_i = 0|X_i$.

Table 4.4: Data and design matrix

Y_i	X_i
Y_1	X_1
Y_2	X_2
$\vdots$	$\vdots$
Y_k	X_k
Y_{k+1}	X_{k+1}
$\vdots$	$\vdots$
Y_n	X_n

$$Y_i = \begin{cases} 0 & \text{if} \quad X_i \leq X^*, \quad \text{i=1, 2, \ldots k} \\ 1 & \text{if} \quad X_i > X^*, \quad \text{i=k+1, k+2, \ldots n} \end{cases} \tag{4.4}$$

or

$$Y_i = \begin{cases} 1 & \text{if} \quad X_i \leq X^*, \quad \text{i=1, 2, \ldots k} \\ 0 & \text{if} \quad X_i > X^*, \quad \text{i=k+1, k+2, \ldots n} \end{cases} \tag{4.5}$$

The success of Y_i is a binary response variable which is conditionally independent given X_i. The model can be written as

$$Y_i \mid X_i \sim \text{Bernoulli}(\pi_i), \tag{4.6}$$

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 X_i \quad ,$$

where the probability of success is

$$\pi_i = \text{P}(\text{S}_i) = \text{P}(\text{Y}_i = 1|X_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)} \tag{4.7}$$

and

$$\text{P}(\text{F}_i) = 1 - \text{P}(\text{S}_i)$$

4.4.1 Probability of Complete Separation

The probability of complete separation at X_k for the criterion described in equation (4.4) and (4.5) may be defined as

$$\begin{aligned} \text{P(Complete Separation at } X_k) &= P\Big[\Big(F_1 \bigcap F_2 \bigcap \ldots \bigcap F_k \bigcap S_{k+1} \bigcap \ldots \bigcap S_n\Big) \\ &\quad \bigcup \Big(S_1 \bigcap S_2 \bigcap \ldots \bigcap S_k \bigcap F_{k+1} \bigcap \ldots \bigcap F_n\Big)\Big] \\ &= \prod_{i=1}^{k} P(F_i) \prod_{i=k+1}^{n} P(S_i) + \prod_{i=1}^{k} P(S_i) \prod_{i=k+1}^{n} P(F_i) \end{aligned} \tag{4.8}$$

Now k can be any value between 1 and n-1. Therefore, the probability of complete separation will be

$$\begin{aligned} \text{P(Complete Separation)} &= \prod_{i=1}^{n} P(F_i) + \sum_{k=1}^{n-1} \Bigg[\prod_{i=1}^{k} P(F_i) \prod_{i=k+1}^{n} P(S_i) \\ &\quad + \prod_{i=1}^{k} P(S_i) \prod_{i=k+1}^{n} P(F_i)\Bigg] + \prod_{i=1}^{n} P(S_i) \end{aligned} \tag{4.9}$$

4.4.2 Probability of Quasi-complete Separation

For quasi-complete separation the criterion described in (4.4) should be modified slightly. In the criterion of equality holds for at least one for each of the categories i.e.

$$Y_i = \begin{cases} 0 & & & \text{if} \quad X_i < X^*, & \text{i=1, 2, \ldots k-1} \\ 0 & \text{or} & 1 & \text{if} \quad X_i = X^*, & \text{i=k, k+1} \\ 1 & & & \text{if} \quad X_i > X^*, & \text{i=k+2, k+3, \ldots n} \end{cases} \tag{4.10}$$

or

$$Y_i = \begin{cases} 1 & & & \text{if} \quad X_i < X^*, & \text{i=1, 2, \ldots k-1} \\ 1 & \text{or} & 0 & \text{if} \quad X_i = X^*, & \text{i=k, k+1} \\ 0 & & & \text{if} \quad X_i > X^*, & \text{i=k+2, k+3, \ldots n} \end{cases} \tag{4.11}$$

Now we define the probability of quasi-complete separation (QCS)at X_k as

$$\begin{aligned}
\text{P(QCS at } X_k) &= P\Big[\Big(F_1 \bigcap F_2 \bigcap \ldots \bigcap F_k \bigcap S_{k+1} \bigcap \ldots \bigcap S_n\Big) \bigcup \\
&\quad \Big(F_1 \bigcap F_2 \bigcap \ldots \bigcap S_k \bigcap F_{k+1} \bigcap S_{k+2} \bigcap \ldots \bigcap S_n\Big) \bigcup \\
&\quad \Big(S_1 \bigcap S_2 \bigcap \ldots \bigcap S_{k-1} \bigcap F_k \bigcap S_{k+1} \bigcap F_{k+2} \bigcap \ldots \bigcap F_n\Big) \\
&\quad \bigcup \Big(S_1 \bigcap S_2 \bigcap \ldots \bigcap S_k \bigcap F_{k+1} \bigcap \ldots \bigcap F_n\Big)\Big] \qquad (4.12) \\
&= \prod_{i=1}^{k} P(F_i) \prod_{i=k+1}^{n} P(S_i) + \prod_{i=1}^{k-1} P(F_i)P(S_k)P(F_{k+1}) \\
&\quad \prod_{i=k+2}^{n} P(S_i) + \prod_{i=1}^{k-1} P(S_i)P(F_k)P(S_{k+1}) \prod_{i=k+2}^{n} P(F_i) \\
&\quad + \prod_{i=1}^{k} P(S_i) \prod_{i=k+1}^{n} P(F_i) \qquad (4.13)
\end{aligned}$$

$$\begin{aligned}
&= \{P(F_k)P(S_{k+1}) + P(S_k)P(F_{k+1})\} \\
&\quad \left\{\prod_{i=1}^{k-1} P(F_i) \prod_{i=k+2}^{n} P(S_i) + \prod_{i=1}^{k-1} P(S_i) \prod_{i=k+2}^{n} P(F_i)\right\} \qquad (4.14)
\end{aligned}$$

As k can take any value between 1 and $n-1$ the probability of complete separation would be

$$\begin{aligned}
\text{P(QCS)} &= \left[\{P(F_1)P(S_2) + P(S_1)P(F_2)\} \left\{\prod_{i=3}^{n} P(S_i) + \prod_{i=3}^{n} P(F_i)\right\}\right] \\
&\quad + \sum_{k=2}^{n-2} [\{P(F_k)P(S_{k+1}) + P(S_k)P(F_{k+1})\} \\
&\quad \left\{\prod_{i=1}^{k-1} P(F_i) \prod_{i=k+2}^{n} P(S_i) + \prod_{i=1}^{k-1} P(S_i) \prod_{i=k+2}^{n} P(F_i)\right\}\Bigg] \\
&\quad + [\{P(F_{n-1})P(S_n) + P(S_{n-1})P(F_n)\} \\
&\quad \left\{\prod_{i=3}^{n-2} P(F_i) + \prod_{i=3}^{n-2} P(S_i)\right\}\Bigg] \qquad (4.15)
\end{aligned}$$

4.4.3 Theorem

In this section we introduce a theorem that will be useful in the context of devising an optimality criterion to handle separation problem at the design stage.

Already we know that separation causes non-existence of likelihood estimates in logit models. We want to minimize probability of separation and thereby maximize the probability of existence of likelihood estimates in the experiments. We will devise probability-based optimality criteria that will handle the separation problem. Probabilities of complete and quasi-complete separations are two distinct but related terms. To minimize probability of separation, it is sufficient to deal with complete separation only rather than considering quasi-complete separately. Also, we will deal with complete separation as it is difficult to handle if it happens in real life. The following theorem will prove that if separation occurs at a point where two values of a covariate are equal, then a small changes from that point in either directions (positive or negative) will reduce the probability of separation.

Theorem 4.4.1. *Given a design $X_1 \leq X_2 \leq \ldots \leq X_n$ with $X_k = X_{k+1} = X^*$ for some k, the probability of separation under logistic model is reduced by changing X_k and X_{k+1} to ${X_k}^* = X^* - \delta$ and $X^*_{k+1} = X_{k+1} + \delta$ for a small $\delta > 0$*

Proof. If separation happens at X_k it could belong to either complete or quasi-complete separation. Therefore, we consider both complete and quasi-complete separations separately with regard to this proof.

Complete Separation

We consider the ith response as before

$$Y_i = \begin{cases} 0 & \text{if} \quad X_i \leq X^*, \quad i = 1, 2, \ldots k+1 \\ 1 & \text{if} \quad X_i > X^*, \quad i = k+2, \ldots n \end{cases} \tag{4.16}$$

or

$$Y_i = \begin{cases} 1 & \text{if} \quad X_i \leq X^*, \quad i = 1, 2, \ldots k+1 \\ 0 & \text{if} \quad X_i > X^*, \quad i = k+2, \ldots n \end{cases} \tag{4.17}$$

The corresponding probability of separation is obtained from equation (4.8) as

$$\begin{aligned} \text{P(Complete Separation at X}_k) &= \prod_{i=1}^{k+1} P(F_i) \prod_{i=k+2}^{n} P(S_i) + \\ & \prod_{i=1}^{k+1} P(S_i) \prod_{i=k+2}^{n} P(F_i) \end{aligned} \tag{4.18}$$

where

$$\begin{aligned} \text{P(S}_i|X_i) &= \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)} \\ &= \frac{1}{1 + \exp[-(\beta_0 + \beta_1 X_i)]}; \quad i = 1, 2, \ldots, n \end{aligned} \tag{4.19}$$

and

$$\begin{aligned} \text{P(F}_i|X_i) &= 1 - \text{P(S}_i|X_i) \\ &= \frac{1}{1 + \exp(\beta_0 + \beta_1 X_i)}; \quad i = 1, 2, \ldots, n \end{aligned} \tag{4.20}$$

We illustrate the notations in terms of logistic regression but most of the results should apply to the probit and many other similar models for binary or multinomial regressions as hinted by Albert and Anderson [1984].

Assume $1 + \exp[-(\beta_0 + \beta_1 X_i)] = a_i$ and $1 + \exp(\beta_0 + \beta_1 X_i) = b_i$. Then (4.18) becomes

$$\begin{aligned} \text{P(Complete Separation at X}_k) &= \frac{1}{b_1 b_2 \ldots b_k b_{k+1} a_{k+2} \ldots a_n} \\ &+ \frac{1}{a_1 a_2 \ldots a_k a_{k+1} b_{k+2} \ldots b_n} \end{aligned} \tag{4.21}$$

The expressions for a_k, b_k, a_{k+1}, and b_{k+1} are $a_k = 1 + \exp[-(\beta_0 + \beta_1 X_k)]$, $b_k = 1 + \exp(\beta_0 + \beta_1 X_k)$, $a_{k+1} = 1 + \exp[-(\beta_0 + \beta_1 X_{k+1})]$, and $b_{k+1} = 1 + \exp(\beta_0 + \beta_1 X_{k+1})$ respectively. Now as X_k and X_{k+1} are changed to ${X_k}^* = X^k - \delta$ and $X^*_{k+1} = X_{k+1} + \delta$ the new expressions will be as follows

$a^*_k = 1 + \exp[-(\beta_0 + \beta_1(X_k - \delta))]$, $b^*_k = 1 + \exp(\beta_0 + \beta_1(X_k - \delta))$, $a^*_{k+1} = 1 + \exp[-(\beta_0 + \beta_1(X_k + \delta))]$, and $b^*_{k+1} = 1 + \exp(\beta_0 + \beta_1(X_k + \delta))$

Therefore, the separation probability of (4.21) becomes

$$\text{P(Complete Separation at X}_k)^* = \frac{1}{b_1 b_2 \ldots b_k^* b_{k+1}^* a_{k+2} \ldots a_n} + \frac{1}{a_1 a_2 \ldots a_k^* a_{k+1}^* b_{k+2} \ldots b_n} \tag{4.22}$$

Further, $a_k a_{k+1} = 1 + 2\exp[-(\beta_0 + \beta_1 x_k)] + \exp[-2(\beta_0 + \beta_1 x_k)]$ and $a_k^* a_{k+1}^* = 1+\exp[-(\beta_0+\beta_1 x_k)][\exp(\beta_1\delta)+\exp(-\beta_1\delta)]+\exp[-2(\beta_0+\beta_1 x_k)]$. Since $[\exp(\beta_1\delta)+\exp(-\beta_1\delta)] > 2$ for any β_0, β_1 and $\delta > 0$, $a_k a_{k+1} < a_k^* a_{k+1}^*$ and similarly $b_k b_{k+1} < b_k^* b_{k+1}^*$. Therefore, $\frac{1}{a_k a_{k+1}} > \frac{1}{a_k^* a_{k+1}^*}$ and $\frac{1}{b_k b_{k+1}} > \frac{1}{b_k^* b_{k+1}^*}$ which leads to the conclusion from (4.21) and (4.22) that P(Complete Separation at X_k)>P(Complete Separation at X_k^*).

Quasi-complete Separation

Complete separation is a special case of quasi-complete separation with the vanishing middle terms of (4.13). Comparing equations (4.8) and (4.13) we find the additional terms $\prod_{i=1}^{k-1} P(F_i)P(S_k)P(F_{k+1}) \prod_{i=k+2}^{n} P(S_i) + \prod_{i=1}^{k-1} P(S_i)P(F_k)P(S_{k+1}) \prod_{i=k+2}^{n} P(F_i)$ in (4.13). Therefore, it is evident that probability of complete separation is less than the probability of quasi-complete separation for a particular design. When X_k and X_{k+1} changes to $X_k{}^* = X^* - \delta$ and $X_{k+1}^* = X_{k+1} + \delta$ for a small $\delta > 0$ the problem of quasi-complete separation turns to a complete separation problem. Thus the probability of separation is reduced for quasi-separation problem when X_k and X_{k+1} changes to $X_k{}^* = X^* - \delta$ and $X_{k+1}^* = X_{k+1} + \delta$ for a small $\delta > 0$. Numerical illustrations in this regard are given in section 4.4.4. □

4.4.4 Numerical Example: Reduction of Probability of Separation with the Minor Changes in Design Points

This section explains the theorem proved in Section 4.4.3 by numerical examples. Consider Tables 4.5 and 4.6 as illustrations. In Table 4.5 we find that the probability of complete separation is reduced to 0.11213 from 0.11220 when $X_i = 0$ is changed a bit in both directions by 0.05 unit. Whereas the Table 4.6 shows that the probability of separation reduced from 0.20783 to 0.10495 with a minor change of X_i by 0.01 unit in both directions. Small R programme given in the Appendix

Table 4.5: Probability of complete separation at $X_3 = X^*$

Y_i	X_i	X_i^*.
0	-1	-1
0	0	-0.05
0	0	0.05
1	1	1
1	2	2
1	3	3
P(CS)	0.11220	0.11213

Table 4.6: Probability complete and quasi-complete separations at $X_3 = X^*$

Y_i	X_i	X_i^*.
0	-2	-2
0	-1	-1
0	0	-0.01
1	0	0.01
1	1	1
1	2	2
P(CS)	-	0.10495
P(QCS)	0.20783	-

are used to compute probabilities of complete and quasi-complete separations.

4.5 Probability of Quasi-complete Separation with Two or Three Equal Design Points Successively

We have derived the expression to compute probability of quasi-complete separation when there are two equal successive design points in Section 4.4.2. Therefore, we can extract the formula for quasi-complete separation probability with two equal design points from (4.14). Now we define the probability of quasi-complete

Table 4.7: Separation with two equal successive design points

Y	X
0	-3
0	-2
0	-1
0	0
1	0
1	1
1	2
1	3

separation at X_k as

$$\begin{aligned} \text{P(Quasi-complete Separation at } X_k) &= \{P(F_k)P(S_{k+1}) + P(S_k)P(F_{k+1})\} \\ &\quad \left\{\prod_{i=1}^{k-1} P(F_i) \prod_{i=k+2}^{n} P(S_i) + \prod_{i=1}^{k-1} P(S_i) \right. \\ &\quad \left. \prod_{i=k+2}^{n} P(F_i)\right\} \end{aligned} \tag{4.23}$$

We consider a hypothetical example with a sample size 8 and a response variable Y given in Table 4.7. In this example, there are two equal design points (X=0) for which the response can be 0 or 1.

We can compute probability of separation for the data set given in Table 4.7 using formula (4.23). If separation happens between X_4 and X_5 this will lead to quasi-complete separation, but all other pairs will lead to complete separation problem. There may have some data sets with three or more equal design points. For instance, if $x_k = x_{k+1} = x_{k+2} = x^*$, then probability of quasi-complete separation at x_k is

$$\begin{aligned} \text{P(QCS at } x_k) &= \{[P(F_k)P(F_{k+1})P(S_{k+2}) + P(F_k)P(S_{k+1})P(F_{k+2}) + P(F_k)P(S_{k+1})P(S_{k+2})] \\ &\quad + [P(S_k)P(S_{k+1})P(F_{k+2}) + P(S_k)P(F_{k+1})P(F_{k+2}) + P(S_k)P(F_{k+1})P(S_{k+2})]\} \\ &\quad \left\{\prod_{i=1}^{k-2} P(F_i) \prod_{i=k+3}^{n} P(S_i) + \prod_{i=1}^{k-1} P(S_i) \prod_{i=k+3}^{n} P(F_i)\right\} \end{aligned} \tag{4.24}$$

4.6 Sequential Method to Compute Probability of Separation

In a data set all design points may not be distinct. There may have several sets of two or more equal design points. In that case, the probability formula for complete or quasi-complete separation would not be useful individually. We propose a sequential compound method to compute probability of separation. The formulas of complete and quasi-complete separation probability would be used simultaneously for a single data set.

As an illustration we refer to the Table 4.8 where first column represents assumption on cut off points regarding separation i.e. where cut off points, those might separate responses, may lie. The second column represents the probability of complete or quasi-complete separation on the basis of the assumption described in the first column

Table 4.8: Compound method to compute separation probability

Assumption	Component in the current stage
$x_1 < x^* < x_2$	$Z_1 = \mathrm{P(CS)}$
$x_2 < x^* < x_3$	$Z_2 = \mathrm{P(CS)}$
$x_3 < x^* < x_4$	$Z_3 = \mathrm{P(CS)}$
$x_5 = x_6 = x^*$	$Z_4 = \mathrm{P(QCS)}$
$x_6 < x^* < x_7$	$Z_5 = \mathrm{P(CS)}$
$x_8 = x_9 = x_{10} = x^*$	$Z_6 = \mathrm{P(QCS)}$
...	...
$x_r = x_{r+1} = x_{r+2} = x^*$	$Z_{(r+2)-\sum a_i - 1} = \mathrm{P(QCS)}$
...	...
$x_{n-1} < x^* < x_n$	$Z_{n-\sum a_i - 1} = \mathrm{P(CS)}$

We have described the components in a sequential method to compute probability of separation in Table 4.8 where Z_1 stores the result of probability of complete separation if x^* lies between x_1 and x_2, Z_4 stores the result of probability of quasi-complete separation if $x^* = x_5 = x_6$, n is the total number of design points, for distinct or equality cases each row should contain two design points, thus we define $a_i = d_i - 2$ is the number of additional design points in ith row, d_i is the number of design points in d_i is the number of design points in ith row. Then adding Z_1 ,

$Z_2 \ldots , Z_{n-\sum a_i - 1}$, we have the probability of separation as in equation (4.25).

$$\mathrm{P}(\text{Separation}) = Z_1 + Z_2 + Z_3 + \ldots + Z_{n-\sum a_i - 1} \tag{4.25}$$

4.7 Probability-based Optimality Criteria

All criteria namely D-, G-, E-, A- and so on have been developed either for estimating parameters precisely, or predicting outcome variables efficiently, or minimizing squared length of the largest axis of confidence ellipsoid but not increasing or decreasing the probability of any events. Only probability-based optimality criterion considers a particular aspect of increasing or decreasing an event probability. McGree and Eccleston [2008] have developed a probability-based optimality criterion that considers increasing probability of success of a particular event. As the separation leads to non-existence of maximum likelihood estimates, we want to minimize the probability of separation in the studies to increase the possibility of maximum likelihood estimates. Therefore, we propose a probability-based criterion that will be used to choose such designs which will reduce the probability of separation in the experimental studies and thereby increasing the probability of the ML parameters being estimable.

4.7.1 P_s-optimality Criterion

The general form of the proposed probability-based optimality criterion may be expressed as

$$\begin{aligned} \psi^{P_s} &= f\left(\mathrm{P}(\mathrm{CS}) \text{ and/or } \mathrm{P}(\mathrm{QCS})\right) && (4.26)\\ &= f\left(\mathrm{P}(\mathrm{F}_1), \mathrm{P}(\mathrm{F}_2), \ldots, \mathrm{P}(\mathrm{F}_n)\mathrm{P}(\mathrm{S}_1), \mathrm{P}(\mathrm{S}_2), \ldots, \mathrm{P}(\mathrm{S}_n)\right) && (4.27) \end{aligned}$$

where

$$\mathrm{P}(\mathrm{S}_i) = \mathrm{P}(\mathrm{Y}_i = 1|X_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)} \quad \text{and} \quad \mathrm{P}(\mathrm{F}_i) = 1 - \mathrm{P}(\mathrm{S}_i) \tag{4.28}$$

4.7.2 Compound Criteria and DP$_s$-optimality

A single purpose design may be quite inefficient for handling a real-life problem. Therefore, we often need to incorporate more than one design criterion and a common approach is simply to construct a weighted average, which may depend upon different information matrices. Designs based upon this method have been termed compound designs [Müller and Stehlík, 2010]. However, the concept of compound designs has been described in earlier literature, for instance, Biedermann et al. [2007] implemented (local) compound optimal designs for estimating multiple percentiles on the dose response curve where the design space was restricted due to ethical concerns over drug toxicity and/or efficacy. Atkinson [2008] has combined the notions of parameter estimation (i.e. D-optimality criterion) and model discrimination (i.e. T-optimality criterion which oppose the D-optimality) through a compound criterion DT-optimality to form designs efficient under both criteria. McGree et al. [2008] devise a new class of compound criteria for generalized linear models that offer a method of achieving designs that posses the properties of efficient parameter estimation and high probability of a desired outcome. They formed the probability based compound optimality criterion by maximizing a weighted product of efficiencies.

In our study, it is necessary to estimate parameters precisely along with minimal separation problem. To achieve the dual goals, a compound criterion has been developed considering both D- and P$_s$- optimality. The proposed compound criterion may be expressed as a ratio of D- to P$_s$-optimality, weighted by a pre-defined mixing constant $0 \leq \alpha \leq 1$. Thus we define

$$\mathrm{DP}_s(\xi) = \frac{[P(S|X)]^{(1-\alpha)}}{[|X'WX|]^{\alpha/q}} \tag{4.29}$$

where $P(S|X) \equiv$ P$_s$-optimality which may be considered here as P(Complete Separation) defined in (4.9); W is diag $(w_1 S_1(1-S_1), \ldots, w_n S_n(1-S_n)$ and $\sum w_i = 1$. The mixing constant α balances between D- and P$_s$-optimalities and q is the number of parameters in the model.

We want to minimize (4.29) by choosing optimal design points given some specified values of the parameters. DP$_s(\xi)$ will be minimum when the probability of separa-

tion is minimum (i.e. P_s-optimality is close to zero) and $[|X'WX|]^{\alpha/q}$ is maximum (i.e. D-criterion is low). The simultaneous effect of D- and P_s-optimality criteria will determine the value of DP_s-optimality criterion.

Again D- and P-efficiencies can be defined as

$$D_{\text{eff}}(\xi) = \left(\frac{|M(\xi)|}{|M(\xi_D^*)|}\right)^{1/q} \quad \text{and} \quad P_{\text{seff}}(\xi) = \left(\frac{P(\xi_{P_s}^*)}{P(\xi)}\right) \tag{4.30}$$

where $|M(\xi_D^*)|$ is the determinant of the information matrix of a D-optimal design and $|M(\xi)|$ is a determinant of any other design with the same size. $P(\xi_{P_s}^*)$ is the probability of separation for a P_s-optimal design with the same size. However, McGree and Eccleston [2008] derived combined criterion on the basis of D- and P-efficiencies, whereas our criterion is different than theirs in the context that ours should be minimized while their's was maximized. They motivated their DP-optimality approach through an application to a potato packaging experiment. DP-optimality was applied to this application to decrease the probability of observing liquid in the potato pack. Simple D-optimal designs is quite likely to produce potato packs with liquid in them. However, DP-optimality criterion enables designs that satisfy dual goals- parameter estimation and low probability of observing liquid in the potato packs.

In our case, we propose compound DP_s- criterion that enables precise parameter estimation and low probability of separation. Our criterion differs with the criterion of McGree and Eccleston [2008] with respect to the probability component where they minimize the probability of having liquid in the potato packages whereas we minimize the probability of separation. The algebraic forms of probability calculation in two cases are also different. The model of McGree and Eccleston [2008] was derived from a central composite design given in Woods et al. [2006]:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_2 x_3 + \beta_5 x_1^2 + \beta_6 x_2^2 + \beta_7 x_3^2 \tag{4.31}$$

where x_1, x_2, and x_3 denote the vitamin concentration in the pre-packing dip and the levels of two gases in the packing atmosphere respectively. However, our criterion is based on the probability of separation given in equation (4.9) which is combined with D-criterion. Further discussions on compound designs are available

in many papers, for example, Biedermann et al. [2007], Atkinson [2008], McGree et al. [2008], Müller and Stehlík [2010], and McGree et al. [2012].

4.8 Local Optimization of DP_s-optimality Criterion

The equivalence theorem is used to check D-optimality, however, for probability based optimality criterion equivalence theorem is not appropriate. Because the idea of equivalence theorem is based on information matrix which is independent of sample size whereas DP_s-criterion varies with the sample size/design size. Therefore, we look for exact optimal designs for particular sample sizes.

Our proposed compound DP_s-optimal design is a constrained optimal design as values in design should be ordered to make easier the identification of separation. We will deal with constrained optimization of the objective function (4.29) as unconstrained optimization leads to a wrong design with respect to separation problem. For example, Nelder-Mead or simulated annealing method is not suitable for optimization of our DP_s-criterion as Nelder-Mead's simplex or simulated annealing method produces such unordered designs, after ordering the resultant designs give high probability of separation. For example, let us consider a design with size 8. The optimization of DP_s-criterion by Nelder-Mead algorithm provides the design as $X = c($ 32.4999, -5.2183, -9.4666, -58.2227, 17.8507, 42.0957, 16.7786, -12.9247) which after ordering is $X = c$(-58.2227, -12.9247, -9.4666, -5.2183, 16.7786, 17.8507, 32.4999, 42.0957) and the probability of separation is 0.9999206. Again the optimization by simulated annealing method gives the design as $X = c$(-9.4397, 4.9031, 7.2938, 3.3207, 0.7459, -7.5615, -14.4906, -7.3740) which after ordering is $X = c$(-14.4906, -9.4396, -7.5616, -7.3740, 0.7459, 3.3207, 4.9031, 7.2938) and the probability of separation is 0.9677993. However, constrained optimization of DP_s-criterion with $\alpha = 0.5$ gives the design $X = c(-0.1885, -0.1884, -0.1883, -0.1882, -0.1881, -0.1880, 1.6633, 1.6634)$ which provides the probability of separation as 0.0884. By constrained optimization we mean that during optimization we put restriction in the codes in such a way that enables the values of the resultant design to be ordered. Thus, unconstrained optimizations (algorithm that provides unordered values in the design) are

not much useful to find optimal designs to have minimal separation problem.

We have used a R package *optim* to find the optimal designs. The *optim* is concerned with the general-purpose optimization which is based on Nelder-Mead, quasi-Newton, simulated annealing, L-BFGS-B, and conjugate-gradient algorithms. We have used box-constrained optimization algorithm L-BFGS-B for local optimization of our criteria as unconstrained or other algorithms provides such designs that lead to high probability of separation. Further, the designs we obtained through optimization by using the R package *optim* are considered optimal as with different set of starting values in *optim* we ended up with the same designs.

4.9 DP_s-optimal Designs for Models with More Than One Factor

The complete and quasi-complete separation were defined for one factor in Section 4.4. If there are $(p-1)$ factors $X_1, X_2, \ldots, X_{p-1}$ and the parameters in the logistic model are $\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_{p-1}$. For a set of coefficients $c_0, c_1, c_2, \ldots, c_{p-1}$ complete separation is defined as

$$Y_i = \begin{cases} 0 & \text{if} \quad c_0 + c_1 X_{i1} + c_2 X_{i2} + \ldots + c_{p-1} X_{ip-1} \leq M, \quad i = 1, 2, \ldots k \\ 1 & \text{if} \quad c_0 + c_1 X_{i1} + c_2 X_{i2} + \ldots + c_{p-1} X_{ip-1} > M, \quad i = k+1, k+2, \ldots n \end{cases} \tag{4.32}$$

or

$$Y_i = \begin{cases} 1 & \text{if} \quad c_0 + c_1 X_{i1} + c_2 X_{i2} + \ldots + c_{p-1} X_{ip-1} \leq M, \quad i = 1, 2, \ldots k \\ 0 & \text{if} \quad c_0 + c_1 X_{i1} + c_2 X_{i2} + \ldots + c_{p-1} X_{ip-1} > M, \quad i = k+1, k+2, \ldots n \end{cases} \tag{4.33}$$

where M is a cut-off point that separates the responses into two categories. We

can express the design matrix as

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1(p-1)} \\ 1 & X_{21} & X_{22} & \dots & X_{2(p-1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{i1} & X_{i2} & \dots & X_{i(p-1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n(p-1)} \end{bmatrix} \tag{4.34}$$

Now following (4.29) the DP_s-criterion for $(p-1)$ factors can be defined as

$$DP_s(\xi) = \frac{[P(S|\mathbf{X},\beta)]^{(1-\alpha)}}{[|\mathbf{X}'\mathbf{W}\mathbf{X}|]^{\alpha/p}} \tag{4.35}$$

For the given values of $\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_{p-1}$ we need to choose design matrix $\mathbf{X}$ shown in (4.34) in such a way that the quantity in (4.35) is minimum i.e. we need to optimize (4.35). There are several challenges in the phase of optimization of DP_s-criterion for more than one factor. Firstly, it is essential to identify M and all possible linear combinations that characterizes separation problem. Secondly, the optimization itself is challenging as optimization algorithm is needed to be developed to optimize the criterion in case of multi-factor experiment. However, this interesting further extension can be done in future studies.

4.10 Comparing Results from Various Designs

4.10.1 Comparing D- and P_s-optimality from D-, P_s, DP_s-optimal designs

D-optimal designs are widely used in practice. It has particular feature of precise parameter estimation. Table 4.9 compares D- and P_s-optimality along with D- and P-efficiencies from D-optimal designs. About half of the design points are negative and rest of the points are positive in D-optimal designs shown in Table 4.9. D-criterion remains unaltered as expected with even or odd number of sample sizes.

Table 4.9: Comparing D- and P_s-optimality from D-optimal designs

Size	Design	D-criterion	P_s-criterion	D-eff	P-eff
2	-1.5434 1.5434	19.95	1	1	1
3	-1.5434 1.5434 1.5434	22.45	0.8550	1	0.8772
4	-1.5434 -1.5434 1.5434 1.5434	19.95	0.7099	1	0.7043
5	-1.5433 -1.5433 1.5432 1.5432 1.5432	20.78	0.5859	1	0.5334
6	-1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434	19.95	0.4829	1	0.3883
7	-1.5435 -1.5434 -1.5433 1.5433 1.5434 1.5435 1.5436	20.37	0.3980	1	0.2749
8	-1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434	19.95	0.3279	1	0.1906
9	-1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	20.20	0.2702	1	0.1301
10	-1.5434 -1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	19.95	0.2226	1	0.0877
11	-1.5434 -1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	20.12	0.1834	1	0.0586
12	-1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	19.95	0.1511	1	0.0388
13	-1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	20.07	0.1245	1	0.0255
14	-1.5435 -1.5435 -1.5435 -1.5435 -1.5435 -1.5435 -1.5435 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	19.95	0.1026	1	0.0167
15	-1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 1.5435 1.5435 1.5435 1.5435 1.5435 1.5435 1.5435 1.5435	20.04	0.0846	1	0.0108
16	-1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5433 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	19.95	0.0697	1	0.0070
17	-1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	20.02	0.0574	1	0.0045
18	-1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	19.95	0.0473	1	0.0029
19	-1.5436 -1.5436 -1.5436 -1.5436 -1.5436 -1.5436 -1.5436 -1.5436 -1.5421 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	20.01	0.0390	1	0.0019
20	-1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434 1.5434	19.95	0.0321	1	0.0012

Table 4.10: Comparing D- and P_s-optimality from P_s-optimal designs

Size	Design	D-criterion	P_s-criterion	D-eff	P-eff
2	-0.1000 -0.0990	64317470.1860	1	0.0006	1
3	-0.0001 -0.0000 0.0001	2400000010.0000	0.7500	0.0001	1
4	-0.0460 -0.0001 0.0000 0.0001	40327.8124	0.5000	0.0222	1
5	-0.0002 -0.0001 -0.0000 0.0001 0.0002	800000010.8000	0.3125	0.0002	1
6	-0.0441 -0.0002 -0.0001 0.0000 0.0001 0.0002	59053.9891	0.1875	0.0184	1
7	-0.0003 -0.0002 -0.0001 -0.0000 0.0001 0.0002 0.0003	400000011.0000	0.1094	0.0002	1
8	-0.0003 -0.0002 -0.0001 -0.0000 0.0001 0.0002 0.0003 0.0004	304761915.8095	0.0625	0.0003	1
9	-0.0411 -0.0004 -0.0003 -0.0002 -0.0001 0.0000 0.0001 0.0002 0.0003	96093.3242	0.0352	0.0145	1
10	-0.0004 -0.0003 -0.0002 -0.0001 -0.0000 0.0001 0.0002 0.0003 0.0004 0.0005	193939405.0424	0.0195	0.0003	1
11	-0.0410 -0.0004 -0.0003 -0.0002 -0.0001 -0.0000 0.0001 0.0002 0.0003 0.0004 0.0005	115214.2102	0.0107	0.0132	1
12	-0.0006 -0.0005 -0.0004 -0.0003 -0.0002 -0.0001 0.0000 0.0001 0.0002 0.0003 0.0004 0.0005	134265745.3986	0.0059	0.0004	1
13	-0.0007 -0.0006 -0.0005 -0.0004 -0.0003 -0.0002 -0.0001 0.0000 0.0001 0.0002 0.0003 0.0004 0.0005	114285725.6339	0.0032	0.0004	1
14	-0.0007 -0.0006 -0.0005 -0.0004 -0.0003 -0.0002 -0.0001 0.0000 0.0001 0.0002 0.0003 0.0004 0.0005 0.0006	98461549.6134	0.0017	0.0005	1
15	-0.0014 -0.0013 -0.0011 -0.0010 -0.0009 -0.0008 -0.0007 -0.0006 -0.0005 -0.0004 -0.0003 -0.0002 -0.0001 0.0000 0.0001	80023773.9526	0.0009	0.0005	1
16	-0.0399 -0.0007 -0.0006 -0.0005 -0.0004 -0.0003 -0.0002 -0.0001 0.0000 0.0001 0.0002 0.0003 0.0004 0.0005 0.0006 0.0007	171205.3221	0.0005	0.0108	1
17	-0.0537 -0.0182 -0.0001 -0.0000 0.0001 0.0002 0.0003 0.0004 0.0005 0.0006 0.0007 0.0008 0.0009 0.0010 0.0011 0.0012 0.0013	91253.5832	0.0003	0.0148	1
18	-0.0545 -0.0195 -0.0010 -0.0009 -0.0008 -0.0007 -0.0006 -0.0005 -0.0004 -0.0003 -0.0002 -0.0001 0.0000 0.0001 0.0002 0.0003 0.0004 0.0005	95518.5830	0.0001	0.0145	1
19	-0.0631 -0.0322 -0.0110 -0.0005 -0.0004 -0.0003 -0.0002 -0.0001 0.0000 0.0001 0.0002 0.0003 0.0004 0.0005 0.0006 0.0007 0.0008 0.0009 0.0010	66220.2177	0.0001	0.0174	1
20	-0.0689 -0.0414 -0.0205 -0.0067 -0.0001 -0.0000 0.0001 0.0002 0.0003 0.0004 0.0005 0.0006 0.0007 0.0008 0.0009 0.0010 0.0011 0.0012 0.0013 0.0014	52296.7982	0.0000	0.0195	1

Table 4.11: Comparing D- and P_s-optimality from DP_s-optimal designs with $\alpha = 0.5$

Size	Design	D-criterion	P_s-criterion	D-eff	P-eff
2	-1.5434 1.5434	19.95	1	1	1
3	-1.0476 -1.0475 1.7718	23.71	0.8078	0.9730	0.9285
4	-1.1997 -1.1996 1.1996 1.1997	21.95	0.6442	0.9535	0.7762
5	-0.6823 -0.6822 -0.6821 1.4147 1.4148	27.01	0.4318	0.8771	0.7238
6	-0.4084 -0.4083 -0.4082 -0.4081 1.5497 1.5498	33.88	0.2628	0.7675	0.7135
7	-0.2679 -0.2678 -0.2677 -0.2676 -0.2675 1.6226 1.6227	40.56	0.1541	0.7086	0.7098
8	-0.1885 -0.1884 -0.1883 -0.1882 -0.1881 -0.1880 1.6633 1.6634	46.87	0.0884	0.6524	0.7071
9	-0.1397 -0.1396 -0.1395 -0.1394 -0.1393 -0.1392 -0.1391 1.6871 1.6872	52.90	0.0499	0.6180	0.7048
10	-0.1077 -0.1076 -0.1075 -0.1074 -0.1073 -0.1072 -0.1071 -0.1070 1.7015 1.7016	58.73	0.0278	0.5829	0.7027
11	-0.0855 -0.0854 -0.0853 -0.0852 -0.0851 -0.0850 -0.0849 -0.0848 -0.0847 1.7105 1.7106	64.44	0.0153	0.5588	0.7008
12	-0.0696 -0.0695 -0.0694 -0.0693 -0.0692 -0.0691 -0.0690 -0.0689 -0.0688 -0.0687 1.7161 1.7162	70.05	0.0084	0.5337	0.6990
13	-0.0578 -0.0577 -0.0576 -0.0575 -0.0574 -0.0573 -0.0572 -0.0571 -0.0570 -0.0569 -0.0568 1.7197 1.7198	75.59	0.0046	0.5153	0.6975
14	-0.0488 -0.0487 -0.0486 -0.0485 -0.0484 -0.0483 -0.0482 -0.0481 -0.0480 -0.0479 -0.0478 -0.0477 1.7219 1.7220	81.09	0.0025	0.4960	0.6961
15	-0.0418 -0.0417 -0.0416 -0.0415 -0.0414 -0.0413 -0.0412 -0.0411 -0.0410 -0.0409 -0.0408 -0.0407 -0.0406 1.7231 1.7232	86.56	0.0013	0.4812	0.6948
16	-0.0362 -0.0361 -0.0360 -0.0359 -0.0358 -0.0357 -0.0356 -0.0355 -0.0354 -0.0353 -0.0352 -0.0351 -0.0350 -0.0349 1.7238 1.7239	92.00	0.0007	0.4657	0.6937
17	-0.0317 -0.0316 -0.0315 -0.0314 -0.0313 -0.0312 -0.0311 -0.0310 -0.0309 -0.0308 -0.0307 -0.0306 -0.0305 -0.0304 -0.0303 1.7241 1.7242	97.42	0.0004	0.4534	0.6928
18	-0.0280 -0.0279 -0.0278 -0.0277 -0.0276 -0.0275 -0.0274 -0.0273 -0.0272 -0.0271 -0.0270 -0.0269 -0.0268 -0.0267 -0.0266 -0.0265 1.7240 1.7241	102.83	0.0002	0.4405	0.6918
19	-0.0250 -0.0249 -0.0248 -0.0247 -0.0246 -0.0245 -0.0244 -0.0243 -0.0242 -0.0241 -0.0240 -0.0239 -0.0238 -0.0237 -0.0236 -0.0235 -0.0234 1.7237 1.7238	108.22	0.0001	0.4300	0.6912
20	-0.0224 -0.0223 -0.0222 -0.0221 -0.0220 -0.0219 -0.0218 -0.0217 -0.0216 -0.0215 -0.0214 -0.0213 -0.0212 -0.0211 -0.0210 -0.0209 -0.0208 -0.0207 1.7233 1.7234	113.60	0.0001	0.4191	0.6908

Thus, it is evident that sample size does not matter for D-optimal designs. Probability of separation (P_s-criterion) decreases with the increase of sample size. D-optimal designs with adequate size are acceptable for reducing probability of separation. However, small samples in D-optimal designs might cause substantial separation problem in experiments. Experimenters having adequate resources can choose D-optimal designs with a large sample. For instance, if the experimenter has a sample of only 4 units, their design will have chance of 70.99% to have separation problem i.e. it is likely not to have maximum likelihood estimates of the parameters. However, the probability of separation goes below to 5% if he/she chooses sample size of 18. Further in D-optimal designs, say, there are 10 points on the left (i.e. negative values) and 10 points on the right to zero. Now if we spread the points out slightly, then there would be a little damage to D-optimality but would reduce the probability of separation to a certain amount. For example, if we have the design points as X_1=[-1.54,-1.54,1.54,1.54] then probability of separation is 0.7092589 and if we change the points by small amounts e.g. X_2=[-1.539,-1.538,1.538,1.539] then probability of separation becomes 0.7090707 which is slightly lower than before. However, the corresponding D-criterions are 53.16049 and 53.16248 for X_1 and X_2 respectively.

Table 4.10 compares D- and P_s-optimality from P_s-optimal designs. P_s-optimality criterion takes care of only minimizing separation problem but not the precision of the parameter estimation. Though P_s-optimal designs provide low probability of separation, generally these have high values of D-criterion thus leading to less precise parameter estimation. That is why sole consideration of P_s-optimal designs might not be good choice of designs. However, the P_s-criterion is a basic component in developing DP_s-optimality criterion. With the small number of points in P_s-optimal design we can avoid separation problem to a substantial proportion. When sample size is 8 the probability of separation is about 6.25%, it reduces below 1% with sample size 12 (see Table 4.10 where P_s-criterion corresponds the probability of separation).

Table 4.11 compares D- and P_s-optimality from DP_s-optimal designs with mixing constant $\alpha = 0.5$. DP_s-optimal design balances between D- and P_s-optimality. It is expected that the designs obtained on the basis of DP_s-optimality might be ideal designs that take into account both precise parameter estimation and low

probability of separation. In Table 4.11 we consider $\alpha = 0.5$, that implies equal importance on D- and P_s-optimality. If we are happy with 5% chances of separation then the design size 10 would be enough for an experiment. However, in terms of relative P_s-efficiencies, this is not a good design as it is only 0.06% of effective with respect to local P_s-optimal design. Nevertheless, this design performs well as it has the probability of avoiding separation of 97.22%.

4.10.2 Size of D-, P_s-, and DP_s-optimal designs

In general increasing number of sample points in any design (D-, P_s-, and DP_s-optimal) keeps decreasing the probability of separation. Heinze and Schemper [2002] have also mentioned that probability of separation depends on sample size. However, the reduction of probability of separation in D-optimal design is slower than that of others with the increasing number of sample points (see Figure 4.2). The reduction of probability of separation is faster in P_s-optimal designs, however, these designs are the worst in terms of precision of parameter estimation as the magnitude of the D-criterion is unreasonably high (see Table 4.10). DP_s-optimal designs balance both probability of separation and precision of parameter estimates. If experimenters have adequate resources, they should choose a D-optimal design with large number of sample points, otherwise DP_s-optimal designs would be the best choice, no need to mention that solely P_s-optimal designs are the least favourite choice as there is no guarantee about inference of parameter estimates.

It is evident from the Figure 4.2 that P_s-optimal designs perform better among the comparing optimal designs. However, if the number of sample points exceed 10 both P_s-, and DP_s-optimal designs produce similar probability of separation. Therefore, it is ideal to choose DP_s-optimal designs if the experimenter has resources to afford the number of sample points more than 10.

The magnitude of D-criterion is unrealistically high for P_s-optimal designs shown in Figure 4.3. D-optimal designs produce the minimum values of D-criterion. However, DP_s-optimal designs provide moderate values of D-criterion.

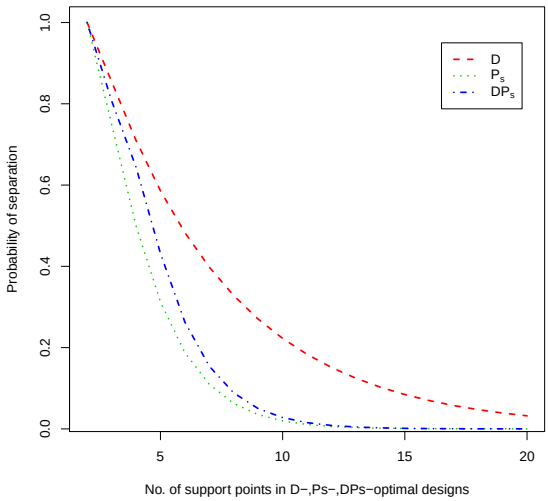

Figure 4.2: Probability of separation in D-, P_s-, and DP_s-optimal designs

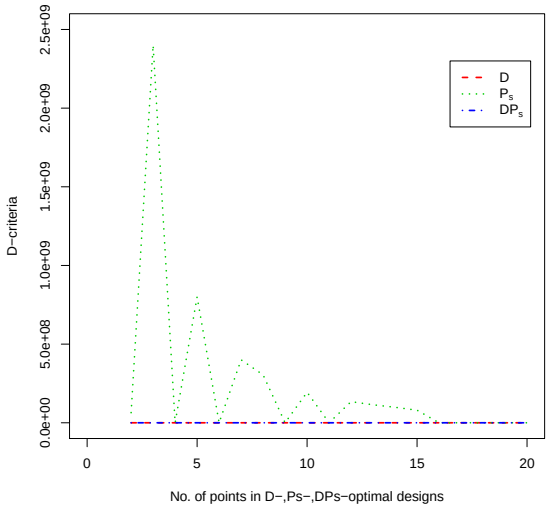

Figure 4.3: D-criterion in D-, P_s, and DP_s-optimal designs

4.10.3 D-and P-efficiencies of D-, P_s, and DP_s-optimal designs

Figure 4.4 compares D- and P-efficiencies together with probability of separation for D-optimal designs. D-efficiency is constant for all D-optimal designs, though probability of separation is not same for all D-optimal designs. Probability of separation decreases with the increased size of D-optimal designs. However, P-efficiency goes down with increasing sample size.

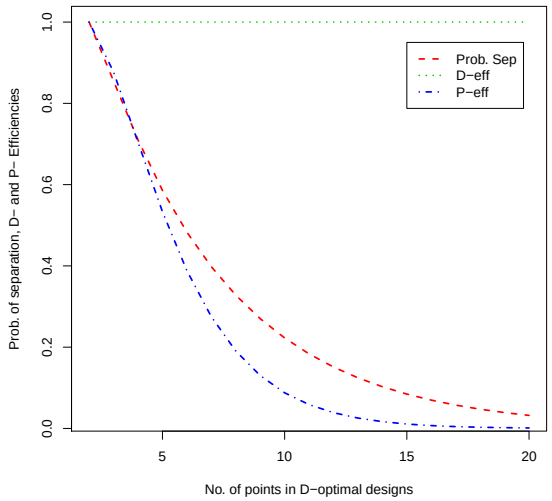

Figure 4.4: D-and P-efficiencies in D-optimal designs with various sizes

As we have seen in Figure 4.3 that the D-criterion was unrealistically high for P_s-optimal design, this leads D-efficiency to be close to zero in Figure 4.5. The probability of separation is in decreasing trend as usual with the rise of number of points. P-efficiency remains constant for any design size.

Figure 4.6 compares D- and P-efficiencies along with probability of separation for DP_s-optimal designs. D-efficiency gradually declines with the higher number of points though P-efficiency remains stable relatively from points 5 up to points 20.

Figure 4.7 compares D-efficiencies in three categories of designs. D-efficiencies remain constant to 1 for D-optimal designs. P_s-optimal designs are completely inefficient in terms of D-efficiency, whereas DP_s-optimal designs maintain a balance between D- and P_s-optimal designs. D-efficiency is reduced approximately to 50% when the number of points reaches to 20 for a DP_s-optimal design.

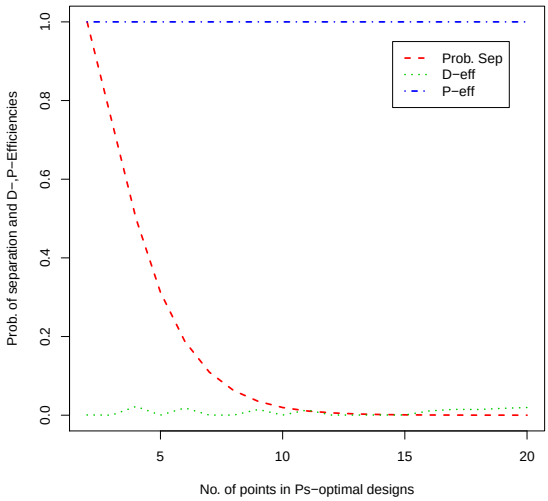

Figure 4.5: D-and P-efficiencies in P_s-optimal designs with various sizes

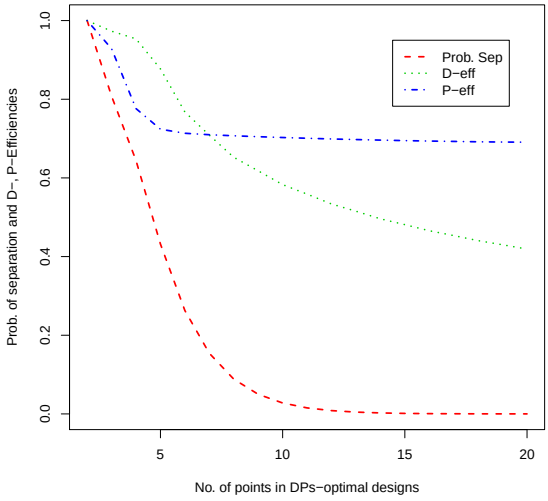

Figure 4.6: D-and P-efficiencies in DP_s-optimal designs with various sizes given that $\alpha = 0.5$

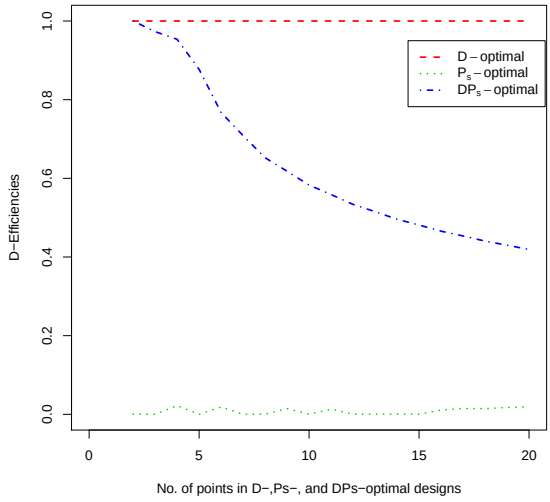

Figure 4.7: D-efficiencies in D-,P_s, and DP_s-optimal designs

P-efficiency for P_s-optimal design is 1 which is ideal. P-efficiencies are decreasing until points of 4 for DP_s-optimal designs shown in Figure 4.8. Then P-efficiencies are decreasing steadily to reach 0.70 around number of points 20. Though at the beginning P-efficiencies are high for D-optimal designs, however, they are not worthwhile for practical use as with low number of design points, the probability of separation is quite high in D-optimal designs. Therefore, higher P-efficiencies in D-optimal designs are not useful information as they are not indicating any reduction of separation problem.

4.10.4 DP_s-optimal designs with different choice of mixing constant α

In Table 4.12 we consider designs with size 8 and 16 to distinguish the results of D- and P_s-optimality. Here α=0 and 1 implies DP_s-optimal design is nothing but P_s- and D-optimal design respectively. With α=0.75 or 1 the designs are D-efficient, however not P_s-efficient and the corresponding probability of separation is around at least 20% and with the minimal probability of separation the designs are not D-efficient. Though the relative P_s-efficiency of the design corresponding to size 16 and α=0.75 is 0.2558, but it could be a desirable design as it provides the chances of separation as low as desired (0.19%), whereas the parameter estimation might be precise also. However, the designs of size 8 with any α might not be a desired

Table 4.12: Comparing D- and P_s-optimal designs with different α

α	Size	Design	D-criterion	P_s-criterion	D-eff	P-eff
0.00	8	-0.0003 -0.0002 -0.0001 -0.0000 0.0001 0.0002 0.0003 0.0004	304761915.81	0.0625	0.0000	1
0.25	8	-0.0390 -0.0389 -0.0388 -0.0387 -0.0386 -0.0385 -0.0384 2.3720	80.66	0.0627	0.3602	0.9971
0.50	8	-0.1885 -0.1884 -0.1883 -0.1882 -0.1881 -0.1880 1.6633 1.6634	46.87	0.0884	0.4664	0.7071
0.75	8	-1.0921 -1.0920 -1.0919 -1.0918 1.0918 1.0919 1.0920 1.0921	23.70	0.1985	0.8418	0.3148
0.90	8	-1.3901 -1.3900 -1.3899 -1.3898 1.3898 1.3899 1.3900 1.3901	20.31	0.2807	0.8559	0.2227
1.00	8	-1.5434 -1.5434 -1.5434 -1.5434 1.5434 1.5434 1.5434 1.5434	19.95	0.3279	1	0.1906
0.00	16	-0.0399 -0.0007 -0.0006 -0.0005 -0.0004 -0.0003 -0.0002 -0.0001 0.0000 0.0001 0.0002 0.0003 0.0004 0.0005 0.0006 0.0007	171205.32	0.0005	0.0108	1
0.25	16	-0.0086 -0.0085 -0.0084 -0.0083 -0.0082 -0.0081 -0.0080 -0.0079 -0.0078 -0.0077 -0.0076 -0.0075 -0.0074 -0.0073 -0.0072 2.3906	154.40	0.0005	0.3595	0.9980
0.50	16	-0.0362 -0.0361 -0.0360 -0.0359 -0.0358 -0.0357 -0.0356 -0.0355 -0.0354 -0.0353 -0.0352 -0.0351 -0.0350 -0.0349 1.7238 1.7239	92.00	0.0007	0.4657	0.6937
0.75	16	-0.1378 -0.1377 -0.1376 -0.1375 -0.1374 -0.1373 -0.1372 -0.1371 -0.1370 -0.1369 -0.1368 -0.1367 1.7364 1.7365 1.7366 1.7367	47.94	0.0019	0.6451	0.2558
0.90	16	-1.1517 -1.1516 -1.1515 -1.1514 -1.1513 -1.1512 -1.1511 -1.1510 1.1510 1.1511 1.1512 1.1513 1.1514 1.1515 1.1516 1.1517	22.64	0.0237	0.8812	0.0211
1.00	16	-1.5441 -1.5441 -1.5431 -1.5431 -1.5431 -1.5431 -1.5425 -1.5425 1.5434 1.5434 1.5439 1.5439 1.5439 1.5439 1.5439 1.5439	19.95	0.0697	1	0.0070

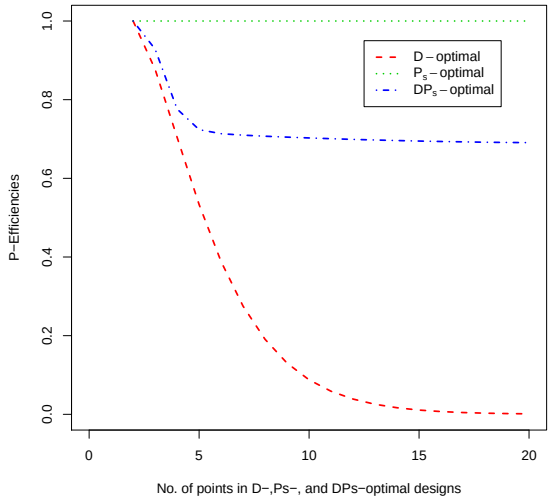

Figure 4.8: P-efficiencies in D-,P$_s$, and DP$_s$-optimal designs

one. Nevertheless, with the scarcity of resources the design of size 8 with α=0.5 might be an optimal choice if any experimenter is happy to take the risk of 8.84% separation probability.

From our empirical experience we see often α=0.90 provides better design (shown in Table 4.12) in terms of P$_s$-optimality, of course, with this choice there should have little compromise with D-optimal condition. In Table 4.12 with the choice of $\alpha = 0.90$ and sample size 16, probability of separation is reduced by roughly 66% while experimenter loses D-efficiency by 11.88%. Thus, experimenter does not lose much as the resultant DP$_s$-optimal design with $\alpha = 0.90$ is slightly D-suboptimal.

The probability of separation goes down with increasing number of sample points. Figure 4.9 shows the change of probability of separation with size of a design for various α. Initially the probability of separation is high for DP$_s$-optimal designs. DP$_s$-optimal designs corresponding to α equal 0 and 1 are nothing but P$_s$- and D-optimal designs respectively. The probability is higher always for DP$_s$- designs with α=1 i.e. for D-optimal designs. However, the probability of separation becomes very small for all designs around the number of sample points 15.

P-efficiencies are high when α is close to zero. The Figure 4.10 shows that P-efficiency is low for the designs with α=1 i.e. for D-optimal designs. P-efficiencies are decreasing with increasing the magnitude of α. P-efficiency is reduced with

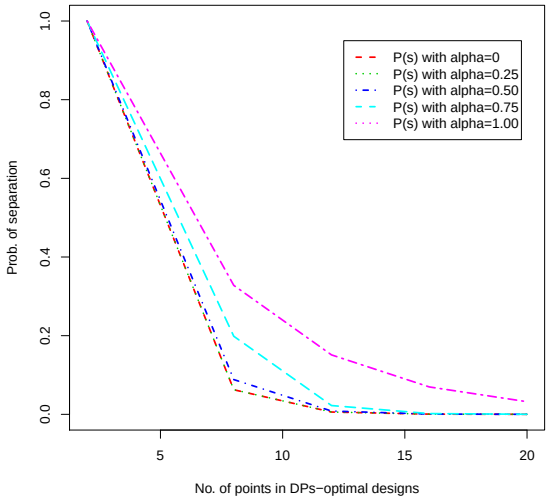

Figure 4.9: Probability of separation Vs size of DP_s-optimal designs with various α

increased number of design points except the designs corresponding to α equal to 0 and 0.25. However, the comparison solely based on P-efficiency without considering probability of separation is not understandable.

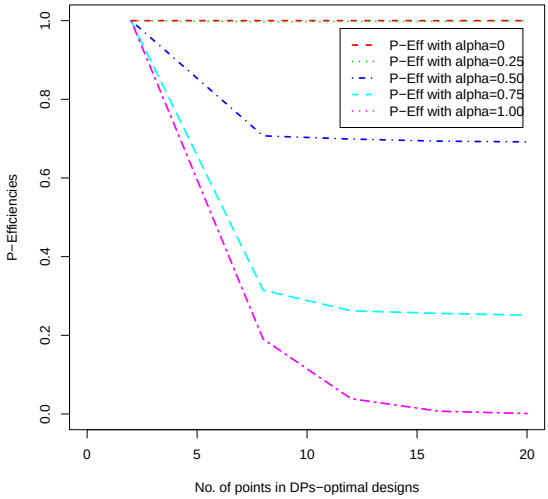

Figure 4.10: P-efficiencies Vs size of DP_s-optimal designs with various α

The values of D-criterion increases with an increase of number of design points shown in Figure 4.11. When there are two design points the D-criterion has the

minimum value for all designs. The value of D-criterion is unrealistically high for the designs with α=0; therefore, were not shown in the figure.

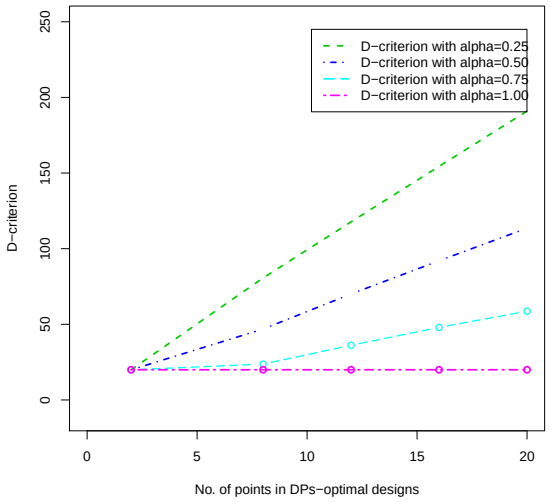

Figure 4.11: D-criterion Vs size of DP$_s$-optimal designs with various α

D-efficiencies are remained almost constant for DP$_s$-optimal designs with α=0 or α=1 shown in Figure 4.12. D-efficiency is the highest for DP$_s$- designs with α=1 i.e. for for D-optimal designs. D-efficiency is the worst for DP$_s$- designs with α=0 i.e. for P$_s$-optimal designs. In general higher the value of α, the higher the value of D-efficiency for a DP$_s$-optimal design. However, D-efficiency decreases with the increase of number of design points.

4.10.5 DP$_s$-optimal designs with different sizes

The Figure 4.13 shows that the probability of separation rises slowly with the increase of mixing constant α though it is roughly stable for designs of all sizes until mixing constant reaches around 0.5. Data is completely separated with design size 2. The probability of separation remains low around α=0.75 for designs with size 12 or more.

The Figure 4.14 portrays D-criterion vs mixing constant for designs with various sizes. It is mentioned earlier that α equal to zero implies a P$_s$-optimal design. The magnitude of D-criterion is unrealistically high for P$_s$-optimal designs; therefore

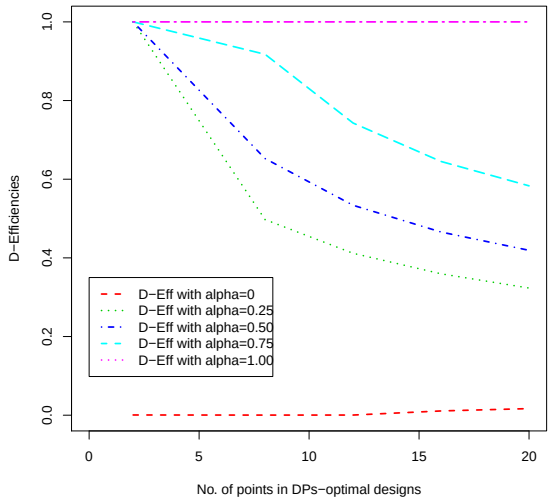

Figure 4.12: D-efficiencies Vs size of DP_s-optimal designs with various α

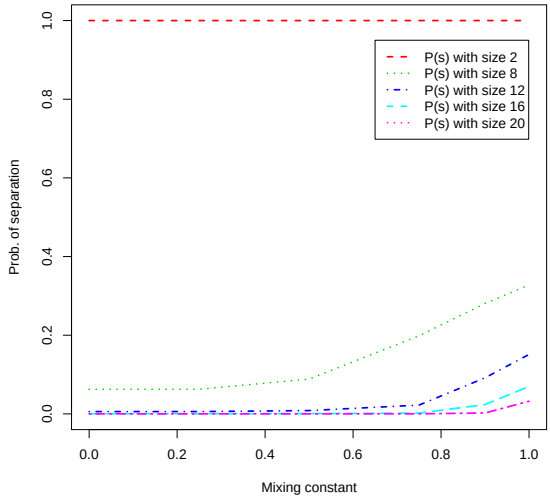

Figure 4.13: Probability of separation Vs mixing constant (α) from DP_s-optimal designs with different sizes

ignored in Figure 4.14. The values of D-criterion are reduced with the rise of mixing constant. Actually the more tendency towards D-optimal designs, the more reduction in the values of D-criterion. D-criterion is low for a design with size 2 and higher for a design with size 20.

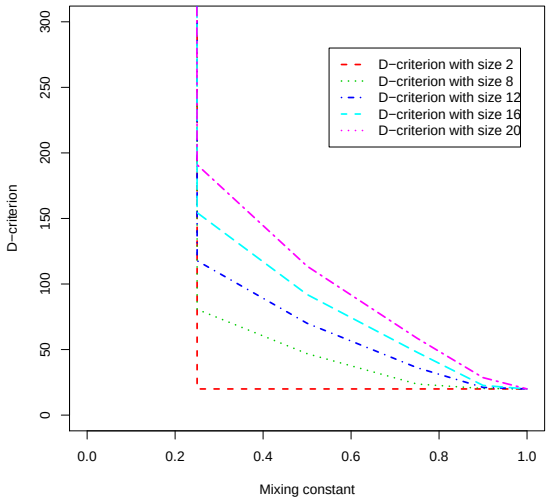

Figure 4.14: D-criterion Vs mixing constant (α) from DP$_s$-optimal designs with different sizes

From Figure 4.15 it seems that the designs with size 2 are P-efficient, however, information is not much useful as probability of separation is high for any designs with size 2 (see Figure 4.13). P-efficiency is reduced steeply for designs of all sizes and is low with mixing constant at 1 i.e. when DP$_s$-optimal designs turns to D-optimal designs.

D-efficiency increases with the rise of mixing constant shown in Figure 4.16. It is close to zero for all designs with α equal to 0 i.e. for P$_s$-optimal designs. On the other hand, D-efficiency is highest for DP$_s$- designs with α equal to 1 i.e. for D-optimal designs.

4.11 Sensitivity Analysis

The importance of sensitivity analysis (SA) of optimality criteria is well recognized as it indicates whether designs will differ radically with modest changes to initial parameter values. It becomes an integral part of the modelling process starting

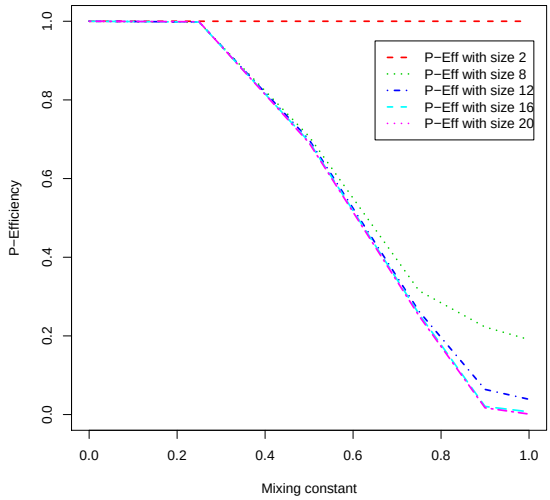

Figure 4.15: P-efficiency Vs mixing constant (α) from DP$_s$-optimal designs with different sizes

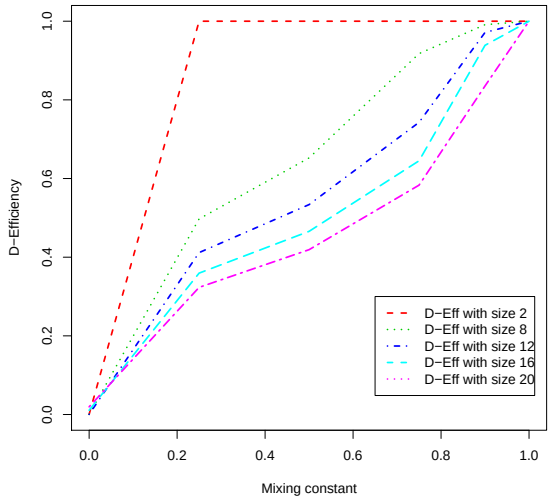

Figure 4.16: D-efficiency Vs mixing constant (α) from DP$_s$-optimal designs with different sizes

from model specification through the analysis and optimization, up to the validation phase [Nestorov, 1999]. The SA is used to quantify the effects of the change of magnitude of model parameters to the outputs of interest (for instance, optimum value of the criterion or change in the probability of separation). Nobody knows the true value of the model structure and/or parameters, however, we use approximations and estimates the latter usually drawn from the small experiment sample. Therefore, it is essential to study how the system will behave if its parameters are not exactly similar as they were assumed.

Let $\mathrm{DP}_s(\xi;\beta_0,\beta_1)$ be our optimality criterion. We have found optimal design by minimizing $\mathrm{DP}_s(\xi;\beta_0,\beta_1)$. We call this design as $\xi^*_{\beta_0,\beta_1}$. The parameter β_0 is changed deliberately a bit by δ and defined in a new position as $\beta_0^* = \beta_0 - \delta$. Now again we find optimal design by minimizing $\mathrm{DP}_s(\xi;\beta_0^*,\beta_1)$. We call this as $\xi^*_{\beta_0^*,\beta_1}$. Now we evaluate

$$\mathrm{S} = \frac{\mathrm{DP}_s(\xi^*_{\beta_0^*,\beta_1};\beta_0^*,\beta_1)}{\mathrm{DP}_s(\xi^*_{\beta_0,\beta_1};\beta_0^*,\beta_1)} \tag{4.36}$$

The quantity S in (4.36) can never exceed 1 as $\mathrm{DP}_s(\xi^*_{\beta_0^*,\beta_1};\beta_0^*,\beta_1)$ is the lowest value with the optimum design $\xi^*_{\beta_0^*,\beta_1}$. Because no other design can be better than $\xi^*_{\beta_0^*,\beta_1}$ given the parameters β_0^* and β_1, Therefore, $\mathrm{DP}_s(\xi^*_{\beta_0,\beta_1};\beta_0^*,\beta_1)$ will be greater than $\mathrm{DP}_s(\xi^*_{\beta_0^*,\beta_1};\beta_0^*,\beta_1)$. However, sensitivity index can be defined as

$$\mathrm{SI} = \frac{|\mathrm{DP}_s(\xi^*_{\beta_0,\beta_1};\beta_0^*,\beta_1) - \mathrm{DP}_s(\xi^*_{\beta_0^*,\beta_1};\beta_0^*,\beta_1)|}{\mathrm{DP}_s(\xi^*_{\beta_0,\beta_1};\beta_0^*,\beta_1)} \times 100 \tag{4.37}$$

The greater the value of SI, the greater the sensitivity of DP_s-optimal due to a change in β_0 parameter value. The indices for the change of β_1 can be defined in the same manner.

Initially the parameter values are considered as β_0=0 and β_1=1. Changes of the order $\Delta\beta_i = \mp 0.5$ are considered, thus the new values of the parameters become β_0^*=-0.5, 0.5 and β_1^*=0.5, 1.5.

Two examples are given in Tables 4.13 and 4.14 with design sizes 16 and 20 respectively. In both Tables α=0 corresponds to the P_s-optimal and α=1 corresponds the D-optimal designs. In both the table it is found that design is affected less for β_1 parameter in P_s-optimal designs and the design is affected less for β_0 parameter in D-optimal designs. We have to be cautious to choose the value of α to obtain a robust design. In Tables 4.13 and 4.14 it is found when $\alpha = 0$, changes in β_1 parameter have little effect on the optimal value of the DP_s-criterion. On the other hand, with greater value of α, say α=0.90, the optimal value of DP_s-criterion is less affected with a minor change in β_0 . When α=0.75 the design is robust for a small change in β_1, however, there is maximum of 28.57% change in the optimum value of the objective function due to small change in the β_0 parameter.

Table 4.13: Sensitivity index of DP_s-optimal designs with design size 16

	$\Delta\beta_0$		$\Delta\beta_1$	
α	-0.5	0.5	-0.5	0.5
0.00	76.44	76.44	0.01	0.01
0.25	65.58	67.39	6.53	3.32
0.50	46.23	40.97	10.55	4.43
0.75	17.10	28.57	13.93	6.77
0.90	2.30	2.26	18.23	9.65
1.00	3.57	3.57	25.48	15.41

Table 4.14: Sensitivity index of DP_s-optimal designs with design size 20

	$\Delta\beta_0$		$\Delta\beta_1$	
α	-0.5	0.5	-0.5	0.5
0.00	87.78	87.63	0.02	0.04
0.25	78.82	80.10	6.53	3.31
0.50	60.65	52.48	10.46	4.34
0.75	27.80	40.62	13.51	06.50
0.90	2.75	8.21	17.76	09.23
1.00	3.57	3.57	25.48	15.41

In some situations, little may be known about the parameters before the experiment is conducted; therefore, initial parameter estimates might not be accurate. In this case, optimality criteria that are sensitive to inaccurate initial parameter estimates may come up with less accurate designs that will have limited practical value.

4.12 Simulation Studies on DP$_s$-optimal Designs

The design performance has been assessed through efficiencies and the sensitivity index in Section 4.11. It is essential to see how robust these designs with the parameter misspecification. The study on robustness with respect to parameter misspecification can be carried out by a simulation study. The simulation study would enable comparison of DP$_s$-designs for different values of α, different sample sizes, and different values of model parameters in a variety of scenarios.

Table 4.15 shows the mean simulated performance of estimates from DP$_s$-designs. The number of simulations has been considered arbitrarily as 10,000 since it was feasible in terms of computing time. The sample size was 8 and 16 and only β_0 varied for these mean based outputs. The summary results that are investigated are bias, standard error (SE) and mean squared error (MSE) of estimates. The bias, SE and MSE are very high in mean based simulated estimates. The main reason is the existence of complete separation or quasi-complete separation in the simulated data set. The quasi- complete or complete separation in some simulation stage causes extreme values of the estimates. Therefore, all the mean based summary measures are unusual i.e. bias, SE, and MSE are very large given various set of true values with different sample sizes. However, in the simulated data set the percentage of separation occurs as expected theoretically. As an illustration, we see that the percentage of complete separation that happens during the simulation studies corresponding to α equals 0.25, 0.5, 0.75, 0.8, 0.9 and 1 are 0.03, 0.05, 0.24, 0.26, 2.57, and 6.95 respectively given that $n = 8$, β_0=0 and β_1=1. The results are similar to the results obtained in theoretical studies (see P$_s$-criterion column in Table 4.12). It is evident that gradually complete separation increases when design moves from P$_s$-optimal designs towards D-optimal designs. However, mean based estimates are not interpretable well because of being very high values of bias, SE, and MSE. Therefore, we will compare and interpret accuracy measures later on the basis of median based information.

Table 4.16 presents the median estimates in the simulated performance of estimators from DP$_s$-designs with sample size 8 and 10,000 simulations. Evidently the medians of point estimates have less biases than previously found mean based estimates (see Table 4.15).

Table 4.15: Simulated performance of mean estimates from DP$_s$ designs with 10,000 simulations

	True Parameter			Bias		SE		MSE		
n	β_0	β_1	α	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	% Sep
			0.25	515.444	13325.170	2308.623	59680.454	5594890.908	3738960616.705	6.88
			0.50	2054.080	16015.440	30375.645	64345.198	926806768.611	4396384764.796	9.00
8	0	1	0.75	392.247	24962.981	87409.981	76063.857	7639894630.373	6408282129.250	20.77
			0.80	1506.063	26316.874	96094.324	75737.334	9235464004.811	6428147964.688	23.28
			0.90	350.624	28096.587	114001.839	77057.408	12995242536.695	6726668603.579	28.01
			1.00	11540.556	18729.058	96489.064	60121.449	9442392917.219	3965004810.814	33.20
			0.25	6766.503	12561.518	31722.479	58889.940	1052000584.732	3625469935.892	7.13
			0.50	10348.026	17058.889	48233.921	65268.935	2433360150.757	4550613590.025	9.72
8	0.5	1	0.75	12209.345	25958.214	96247.385	77104.096	9411700847.403	6618275982.446	20.73
			0.80	13977.620	26718.966	105218.871	76853.936	11265277484.362	6619839965.672	22.69
			0.90	13913.321	28018.533	120544.446	77376.809	14723090838.206	6771610080.509	29.11
			1.00	10840.973	28119.311	129092.823	76573.162	16780817054.155	6653558455.506	33.32
			0.25	4.607	577.536	38.944	4709.996	1537.714	22515395.676	0.03
			0.50	20.652	581.510	207.927	5870.721	43655.978	34800075.353	0.05
16	0	1	0.75	133.868	975.785	1562.298	11390.062	2458450.484	130672686.800	0.24
			0.80	159.459	939.894	1776.778	10375.840	3182051.896	108530689.122	0.26
			0.90	-280.212	2992.273	28688.781	24743.233	823042341.314	621120071.805	2.57
			1.00	-97.290	5708.726	51111.560	32629.016	2612139823.697	1097135742.948	6.95
			0.25	320.256	630.466	3158.080	6217.325	10075033.738	39048748.085	0.05
			0.50	337.306	629.902	3438.029	6421.972	11932640.084	41634377.187	0.06
16	0.5	1	0.75	460.870	794.570	6359.439	9123.848	40650819.465	83867628.263	0.17
			0.80	574.439	1052.158	9660.631	11920.562	93648434.278	143192623.104	0.27
			0.90	1763.727	2420.386	27432.005	21610.560	755550371.000	472827887.085	1.99
			1.00	3432.264	5913.205	56087.807	33583.814	3157307937.823	1162725742.705	7.02

The median estimates indicate that bias and MdMSE would be more precise than that of the estimates based on point estimate mean. The inter-quartile range (IQR) in Table 4.16 describes the difference between 25th and 75th percentiles of the estimates. Under different set of β_0 and β_1 values bias, IQR, and MdSE are roughly minimum given α between 0.75 and 1. This establishes our belief that was claimed in theoretical studies about DP_s-optimal designs on choosing a particular value of α in the experimental studies. Usually IQR declines for $\hat{\beta}_1$ with increase of α and is minimum when α lies between 0.75 and 0.80. Probability of complete separation declines with the reduction of the value of α. MdSE reduces while α increases i.e. when the DP_s-designs tend to D-optimal designs.

Table 4.17 describes the same features in the same pattern as presented in Table 4.16 but all parameters are much lower than that of Table 4.16, that is, bias, IQR, and MdSE (median squared error) are reduced much with the increase of sample size 8 to 16. Probability of complete separation is much lower in comparison to Table 4.16 for all given set of parameter values. It seems that the bias, IQR, and MdSE are minimum when α is around 0.90 for all given set of parameter values. Therefore, a DP_s-design with $\alpha = 0.90$ can be recommended for any study which has problems with separation.

In Table 4.16 and Table 4.17 the true parameters β_0 and β_1 are fixed, but now in Table 4.18 and Table 4.19 the parameter values are coming from some distributions e.g. $\beta_0 \sim N(0, 0.25)$, $\beta_1 \sim N(1, 0.25)$. Table 4.18 and Table 4.19 show the results of simulated performance of median estimates from DP_s-optimal designs with 10000 simulations taking sample sizes 8 and 16. Usually bias reduces with the increase of mixing constant i.e. the more towards D-optimality, the more accuracy of the parameter estimates. This supports the theoretical claims about precise parameter estimability of D-optimal designs. Probability of separation increases with the increase of the value of α i.e. the more towards D-optimality the more probability of separation and probability of separation decreases with the increase of sample size e.g. probability of complete separation is less in Table 4.17 than separation in Table 4.16. For the designs with sample size equal to 16, probability of separation and other indices are better when mixing constant is in between 0.75 and 0.80 i.e. with the designs that are generated by compromising between P_s- and D-optimality.

Table 4.16: Simulated performance of estimates from DP_s designs with 10,000 simulations, n=8

True Parameter			Bias		IQR		MdSE		
β_0	β_1	α	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	% Sep
		0.25	1.052	2.796	238.353	3073.581	1.372	9.074	6.33
		0.50	1.146	2.714	997.121	2650.233	2.090	7.367	9.45
0	0.5	0.75	-0.000	2.105	1.099	10.997	0.302	4.712	20.64
		0.80	0.000	2.181	1.099	10.018	0.302	4.759	22.43
		0.90	-0.000	2.037	2.197	8.442	1.207	4.148	28.03
		1.00	0.001	1.801	2.197	7.602	1.207	3.245	33.26
		0.25	1.041	5.094	119.436	3071.835	1.327	30.403	6.88
		0.50	1.028	5.052	498.256	2650.046	2.090	25.546	9.00
0	1	0.75	0.000	3.892	1.099	21.995	0.302	16.189	20.77
		0.80	0.000	4.074	1.099	19.578	0.302	16.595	23.28
		0.90	-0.000	3.824	2.197	16.883	1.207	14.621	28.01
		1.00	-0.005	3.368	2.197	15.205	1.207	11.341	33.20
		0.25	2.996	2.796	3311.816	3073.581	11.891	9.074	6.67
		0.50	3.588	2.714	3647.103	2649.986	15.399	7.367	9.49
0.5	0.5	0.75	0.003	2.105	9.509	10.998	8.185	4.431	19.97
		0.80	-0.042	2.181	11.026	9.789	16.360	4.759	22.91
		0.90	-0.105	2.037	9.189	8.442	52.461	4.148	29.16
		1.00	0.227	1.801	9.057	7.602	55.631	3.245	32.70
		0.25	2.721	5.094	1655.227	3071.835	10.077	30.404	7.13
		0.50	3.259	5.052	1823.033	2650.046	13.436	25.524	9.72
0.5	1	0.75	0.003	3.892	9.003	21.995	8.824	15.670	20.73
		0.80	-0.042	4.074	9.076	20.036	13.257	16.595	22.69
		0.90	-0.105	3.824	9.935	16.883	45.857	14.621	29.11
		1.00	0.227	3.378	11.057	15.205	53.990	11.412	33.32
		0.25	5.166	5.094	3191.031	3071.835	33.850	30.404	7.12
		0.50	5.501	4.680	3148.427	2650.046	31.549	25.524	9.15
1	1	0.75	0.006	3.892	11.561	21.995	1.095	15.150	20.32
		0.80	-0.083	4.074	11.336	19.578	4.404	16.595	24.06
		0.90	-0.210	3.824	18.302	16.883	72.015	14.621	29.15
		1.00	-0.095	3.378	17.004	15.205	81.666	11.412	34.19
		0.25	4.880	7.508	2126.373	3070.198	30.528	61.524	6.63
		0.50	5.240	6.692	2098.668	2649.859	28.687	52.546	8.82
1	1.5	0.75	0.006	5.757	11.173	32.992	1.095	36.427	20.45
		0.80	-0.083	5.857	10.783	29.366	4.404	34.301	23.31
		0.90	-0.210	5.516	18.245	25.325	66.217	30.432	28.55
		1.00	-0.095	4.870	16.933	22.807	75.730	23.716	33.56

Table 4.17: Simulated performance of estimates from DP$_s$-designs with 10,000 simulations, n=16

True Parameter			Bias		IQR		MdSE		
β_0	β_1	α	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	% Sep
		0.25	0.720	1.770	13.868	877.539	1.089	6.818	0.04
		0.50	0.479	2.314	45.678	653.065	0.554	5.714	0.03
0	0.5	0.75	0.424	1.976	39.390	143.616	0.507	3.903	0.15
		0.80	0.322	0.009	1.699	2.898	0.500	0.195	0.24
		0.90	-0.028	0.039	0.971	0.491	0.300	0.026	1.80
		1.00	0.000	0.130	0.847	1.975	0.179	0.034	6.95
		0.25	0.715	3.248	6.503	815.701	0.967	22.097	0.03
		0.50	0.457	4.218	19.669	541.977	0.517	18.673	0.05
0	1	0.75	0.392	3.663	2.403	7.707	0.449	13.416	0.24
		0.80	0.533	3.548	48.047	279.571	0.529	12.587	0.26
		0.90	-0.000	0.067	0.847	0.991	0.179	0.104	2.57
		1.00	0.000	0.260	0.847	5.265	0.180	0.136	6.95
		0.25	2.094	1.726	765.741	754.158	7.182	6.364	0.03
		0.50	2.514	2.306	664.827	620.942	6.841	5.561	0.08
0.5	0.5	0.75	2.312	2.017	182.916	143.616	5.363	4.068	0.27
		0.80	0.137	0.120	4.348	2.913	0.834	0.195	0.38
		0.90	-0.023	0.033	1.029	0.496	0.342	0.026	2.49
		1.00	0.130	0.130	1.149	1.975	0.379	0.034	6.83
		0.25	1.951	3.250	415.144	815.694	7.064	22.402	0.05
		0.50	2.340	4.233	332.222	620.775	6.064	19.219	0.06
0.5	1	0.75	2.198	3.668	91.824	143.323	4.830	13.452	0.17
		0.80	0.264	0.239	4.238	5.657	0.833	0.780	0.27
		0.90	0.072	0.078	1.116	3.572	0.205	0.110	1.99
		1.00	0.130	0.261	1.149	3.725	0.379	0.137	7.02
		0.25	3.344	3.248	882.848	876.028	23.662	22.380	0.05
		0.50	4.373	4.233	642.944	620.775	19.890	19.219	0.08
1	1	0.75	4.313	3.668	162.793	143.323	18.603	13.492	0.16
		0.80	-0.048	0.239	6.505	5.601	0.889	0.780	0.32
		0.90	-0.046	0.067	1.341	0.991	0.476	0.104	2.15
		1.00	-0.095	0.261	1.286	3.725	0.364	0.137	7.06
		0.25	3.131	4.615	587.551	874.610	20.708	43.931	0.05
		0.50	4.133	6.002	374.448	541.809	17.465	38.763	0.09
1	1.5	0.75	3.901	4.939	8.351	40.056	15.293	24.397	0.20
		0.80	3.988	5.097	218.178	279.302	15.922	25.980	0.21
		0.90	-0.046	0.100	1.341	1.487	0.476	0.234	2.43
		1.00	-0.095	0.391	1.286	5.391	0.448	0.308	7.32

Table 4.18: Simulated performance of estimates from DP_s designs with 10,000 simulations, n=8

True Parameter			Bias		IQR		MdSE		
β_0	β_1	α	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	% Sep
		0.25	0.751	5.094	118.814	3071.835	2.267	30.403	7.92
		0.50	1.017	4.681	498.256	2650.046	2.090	25.546	11.37
$\sim$ N(0, 0.25)	$\sim$ N(1, 0.25)	0.75	0.000	3.892	1.099	21.995	0.302	16.189	21.92
		0.80	0.000	4.074	1.648	20.036	1.207	16.595	24.90
		0.90	-0.000	3.824	1.648	16.883	1.207	14.621	29.73
		1.00	-0.005	3.368	1.648	15.205	1.207	11.341	33.26
		0.25	2.721	5.094	1655.227	3071.835	10.077	30.404	7.73
		0.50	3.259	5.052	1823.468	2650.046	13.436	25.546	11.33
$\sim$ N(0.5, 0.25)	$\sim$ N(1, 0.25)	0.75	0.003	4.024	9.301	21.995	8.824	17.721	21.41
		0.80	-0.042	4.074	11.346	20.036	17.524	16.595	25.11
		0.90	-0.105	3.824	9.584	16.883	45.857	14.621	27.90
		1.00	0.227	3.378	9.057	15.205	53.990	11.412	33.38
		0.25	0.531	3.278	68.864	1784.795	4.094	30.403	32.64
		0.50	0.976	4.186	25.508	1149.196	5.868	30.889	35.05
$\sim$ N(0, 1)	$\sim$ N(1, 1)	0.75	0.000	4.024	9.873	22.498	24.370	20.494	47.75
		0.80	0.000	3.785	9.967	20.494	24.834	20.533	46.44
		0.90	-0.000	3.429	10.115	17.673	25.577	26.826	48.48
		1.00	-0.005	3.023	2.197	15.561	1.207	23.570	49.79
		0.25	2.257	3.564	961.769	1781.809	198.440	33.860	28.77
		0.50	2.974	4.461	25.002	26.534	27.778	28.418	33.22
$\sim$ N(0.5, 1)	$\sim$ N(1, 1)	0.75	0.003	4.024	13.422	22.498	51.841	20.494	39.04
		0.80	0.278	3.785	12.693	20.494	58.685	20.533	39.83
		0.90	0.599	3.429	11.680	17.673	69.502	25.335	46.16
		1.00	3.463	3.378	11.057	15.561	55.625	25.628	45.94

Table 4.19: Simulated performance of estimates DP_s-designs with 10,000 simulations, n=16

True Parameter			Bias		IQR		MdSE		
β_0	β_1	α	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	% Sep
		0.25	0.717	3.248	6.892	815.702	1.074	23.745	0.10
		0.50	0.475	4.218	22.193	620.775	0.567	19.219	0.11
$\sim$ N(0, 0.25)	$\sim$ N(1, 0.25)	0.75	0.392	3.328	2.004	7.303	0.519	11.204	0.29
		0.80	0.512	3.503	25.153	143.372	0.535	12.357	0.23
		0.90	-0.000	0.067	0.847	0.991	0.180	0.142	2.93
		1.00	0.000	0.261	0.973	15.769	0.261	0.229	8.17
		0.25	1.951	3.204	382.837	752.634	7.083	22.402	0.26
		0.50	2.303	4.218	290.313	561.676	5.825	19.137	0.27
$\sim$ N(0.5, 0.25)	$\sim$ N(1, 0.25)	0.75	2.152	3.608	91.221	143.323	4.687	13.110	0.26
		0.80	0.264	0.239	4.086	5.583	0.833	0.845	0.57
		0.90	0.072	0.078	1.116	4.149	0.235	0.256	2.51
		1.00	0.130	0.261	1.288	4.080	0.452	0.229	8.25
		0.25	0.717	2.814	6.430	623.468	3.589	24.129	11.67
		0.50	0.576	3.765	17.217	393.242	2.497	22.623	12.09
$\sim$ N(0, 1)	$\sim$ N(1, 1)	0.75	0.558	0.866	3.737	7.709	1.991	14.129	9.77
		0.80	0.561	2.912	24.850	24.821	1.912	12.826	8.51
		0.90	-0.000	0.067	1.946	6.153	0.947	2.182	14.59
		1.00	0.001	0.261	2.072	16.290	0.947	5.764	23.83
		0.25	2.007	3.048	345.802	676.431	20.103	28.454	11.30
		0.50	2.149	3.946	232.731	441.468	9.119	23.334	10.48
$\sim$ N(0.5, 1)	$\sim$ N(1, 1)	0.75	1.456	3.174	9.114	27.073	6.042	14.520	7.88
		0.80	0.410	0.513	5.269	7.299	4.910	6.316	8.79
		0.90	0.139	0.300	5.866	7.265	2.406	2.784	13.89
		1.00	0.158	2.392	9.278	16.209	15.486	9.134	19.21

Table 4.20: Bayesian designs

n	α	Design values
		Priors: $\beta_0 \sim$ N(0, 0.25) and $\beta_1 \sim$ N(1, 0.25)
8	0.25	-0.0440 -0.0439 -0.0438 -0.0437 -0.0436 -0.0435 -0.0434 2.3420
8	0.50	-0.2014 -0.2013 -0.2012 -0.2011 -0.2010 -0.2009 1.6617 1.6618
8	0.75	-1.0664 -1.0663 -1.0662 -1.0661 1.0721 1.0722 1.0723 1.0724
8	0.80	-1.1574 -1.1573 -1.1572 -1.1571 1.1629 1.1630 1.1631 1.1632
8	0.90	-1.3193 -1.3192 -1.3191 -1.3190 1.3241 1.3242 1.3243 1.3244
8	1	-1.4487 -1.4486 -1.4485 -1.4484 1.4534 1.4535 1.4536 1.4537
16	0.25	-0.0121 -0.0120 -0.0119 -0.0118 -0.0117 -0.0116 -0.0115 -0.0114 -0.0113 -0.0112 -0.0111 -0.0110 -0.0109 -0.0108 -0.0107 2.3710
16	0.50	-0.0514 -0.0513 -0.0512 -0.0511 -0.0510 -0.0509 -0.0508 -0.0507 -0.0506 -0.0505 -0.0504 -0.0503 -0.0502 -0.0501 1.7544 1.7545
16	0.75	-0.1888 -0.1887 -0.1886 -0.1885 -0.1884 -0.1883 -0.1882 -0.1881 -0.1880 -0.1879 -0.1878 -0.1877 1.6561 1.6562 1.6563 1.6564
16	0.80	-0.2857 -0.2856 -0.2855 -0.2854 -0.2853 -0.2852 -0.2851 -0.2850 -0.2849 -0.2848 -0.2847 1.5806 1.5807 1.5808 1.5809 1.5810
16	0.90	-1.1113 -1.1112 -1.1111 -1.1110 -1.1109 -1.1108 -1.1107 -1.1106 1.1031 1.1032 1.1033 1.1034 1.1035 1.1036 1.1037 1.1038
16	1	-1.4585 -1.4584 -1.4583 -1.4582 -1.4581 -1.4580 -1.4560 -1.4559 1.4495 1.4496 1.4497 1.4498 1.4499 1.4500 1.4501 1.4502
		Priors: $\beta_0 \sim$ N(0, 1) and $\beta_1 \sim$ N(1, 1)
8	0.25	-0.2941 -0.2940 -0.2939 -0.2938 -0.2937 0.7492 0.7493 0.7494
8	0.50	-0.4619 -0.4618 -0.4617 -0.4616 -0.4615 1.0285 1.0286 1.0287
8	0.75	-0.9984 -0.9983 -0.9982 0.4592 0.4593 0.4594 0.4595 2.3554
8	0.80	-1.0774 -1.0773 -1.0772 0.2849 0.2850 0.2851 1.0929 2.1963
8	0.90	-1.1530 -1.1529 -1.1528 -0.2407 0.7105 0.7106 0.7107 2.3836
8	1	-2.2636 -0.9499 -0.9498 -0.3535 0.3207 0.9362 0.9363 2.1321
16	0.25	-0.3819 -0.3818 -0.3817 -0.3816 -0.3815 -0.3814 -0.3813 -0.3812 -0.3811 0.5469 0.5470 0.5471 0.5472 0.5473 0.5474 0.5475
16	0.50	-0.4517 -0.4516 -0.4515 -0.4514 -0.4513 -0.4512 -0.4511 -0.4510 -0.4509 0.6647 0.6648 0.6649 0.6650 0.6651 0.6652 0.6653
16	0.75	-0.7634 -0.7633 -0.7632 -0.7631 -0.7630 -0.7629 -0.7628 0.4660 0.4661 0.4662 0.4663 0.4664 0.4665 0.4666 0.4667 3.1402
16	0.80	-0.8043 -0.8042 -0.8041 -0.8040 -0.8039 -0.8038 -0.8037 0.4362 0.4363 0.4364 0.4365 0.4366 0.4367 0.4368 1.3839 3.4688
16	0.90	-0.9322 -0.9321 -0.9320 -0.9319 -0.9318 -0.9317 -0.9316 0.3740 0.3741 0.3742 0.3743 0.3744 1.1771 1.1772 1.1773 4.1337
16	1	-2.7446 -1.2876 -1.0674 -1.0156 -0.9798 -0.9092 -0.3983 0.0196 0.0350 0.5422 0.7934 1.1342 1.1933 1.1934 1.1935 4.2062

4.13 Pseudo Bayesian Designs

The optimal designs with respect to the common optimality criteria depend on the unknown parameter(s). To address the problem Chernoff [1953] suggested to adopt a best guess for the unknown parameter(s), say θ, and termed the resultant designs locally optimal. The main disadvantage of to such an approach is that if unknown parameters are misspecified the resulting optimal designs can be highly inefficient wihin the true setting [Dette et al., 2003]. An alternative robust way is to assume sufficient knowledge about θ to specify a prior distribution for the parameter(s) and to average the respective optimality criteria over the plausible values of θ defined by the prior. This leads to so-called (pseudo) Bayesian optimality criteria [Biedermann et al., 2004]. The reason for naming it pseudo Bayesian as there will be no real data at hand similar to likelihood analysis while designing an experiment. The designs obtained by using the Bayesian criterion are called robust Bayesian designs which are expected to be more robust on parameter misspecification.

We have defined previously DP_s-optimality criterion in equation (4.29) which is a function of β_0 and β_1. Let us define a Bayesian version of the DP_s- criterion such that

$$\mathrm{DP}_{SB} = \int_{\mathbb{R}^2} \mathrm{DP}_S \, \mathrm{P}(\beta_0, \beta_1) \, d\beta_0 \, d\beta_1 \tag{4.38}$$

where $\mathrm{P}(\beta_0, \beta_1)$ is the joint prior distribution of β_0 and β_1. The equation (4.38) is evaluated over the region of $\mathbb{R}^2$. The analytical form of DP_s may not be tractable easily. However, by using MCMC techniques, the Bayesian criterion DP_{SB} can be approximated as

$$\mathrm{DP}_{SB} = \frac{1}{T} \sum_{t=1}^{T} \mathrm{DP}_{st} \tag{4.39}$$

$$\mathrm{DP}_{st} = \frac{\mathrm{P}\left(S|X, \beta_{0t}, \beta_{1t}\right)^{\alpha}}{\left[\acute{X} W X\right]^{\alpha/q}} \tag{4.40}$$

where DP_{st} is the DP_S- criterion at simulation t, $\mathrm{P}\left(S|X, \beta_{0t}, \beta_{1t}\right)$ is the probability of separation with the design matrix X given the coefficients β_{0t} and β_{1t} of β_0 and β_1 at simulation t and relevant priors are $\beta_0 \sim N(\mu_0, \sigma^2_{\beta_0})$ and $\beta_{1t} \sim N(\mu_1, \sigma^2_{\beta_1})$

respectively.

The package *optim* in R can be used to optimize DP_{SB}-criterion and thereby to obtain the Bayesian designs. A few set of pseudo Bayesian designs have been obtained assuming different set of normal priors, for example assuming $\beta_{0t} \sim N(0, 0.25)$ and $\beta_{1t} \sim N(1, 0.25)$ for β_0 and β_1 respectively and the results are presented in Table 4.20.

Usually the range of the values in the Bayesian design is small (when variability in the priors are small) over the design regions than the range of the designs obtained using classical DP_s-criterion and the Bayesian designs are wider than non-Bayesian designs when the variability of priors are comparatively larger with respect to its location. For example, the designs with sizes 8 or 16 with $\alpha = 1$ and priors $\beta_0 \sim N(0, 0.25)$ and $\beta_1 \sim N(1, 0.25)$ have smaller range than the range in non-Bayesian designs (see Table 4.12 and Table 4.20. In general, when the variability in the prior is comparatively large, then the Bayesian designs points are more scattered. The performance of these Bayesian designs will be examined through a simulation study.

A simulation study has been conducted to see the performances of the pseudo Bayesian designs. Table 4.21 shows the simulated performance of median estimates assuming true parameters $\beta_0 = 0$ and $\beta_1 = 1$. The two sets of priors $\beta_0 \sim N(0, 0.25)$, $\beta_1 \sim N(1, 0.25)$ and $\beta_0 \sim N(0, 1)$, $\beta_1 \sim N(1, 1)$ with sample sizes 8 and 16 have been considered. The simulated performance of median estimates reveals that usually the bias, IQR, MdSE and probability of complete separation is less in the Bayesian designs than the DP_s-based non-Bayesian designs. For example, in Table 4.18 when $n = 8$, $\alpha = 0.80$ and $\beta_0 \sim N(0, 1)$, $\beta_1 \sim N(1, 1)$ the proportion of separation is 46.44% whereas in Table 4.21 the corresponding probability of separation is 17.86% with $\alpha = 0.80$ and priors $\beta_0 \sim N(0, 1)$, $\beta_1 \sim N(1, 1)$. The same rows in Tables 4.18 and 4.21 show that bias is less, particularly for β_1, in the Bayesian designs than that of classical designs.

True parameters β_0 and β_1 were fixed in Table 4.21. Simulated performance of median estimates in Bayesian designs have been investigated assuming true parameters coming from normal distributions and the corresponding results are shown

Table 4.21: Simulated performance of estimates in Bayesian designs with 10,000 simulations, True parameters: $\beta_0 = 0, \quad \beta_1 = 1$

	Priors			Bias		IQR		MdSE		
n	β_0	β_1	α	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	% Sep
8	$\sim N(0, 0.25)$	$\sim N(1, 0.25)$	0.25	0.735	5.151	134.275	3071.800	1.405	31.085	7.37
			0.50	1.092	5.019	424.585	2650.048	2.061	25.209	9.99
			0.75	-0.003	3.986	1.099	22.461	0.303	15.891	20.66
			0.80	-0.003	3.929	1.101	20.701	0.303	17.878	22.30
			0.90	-0.002	3.639	1.649	17.756	1.207	13.239	26.54
			1.00	-0.002	3.636	2.197	16.172	1.207	13.217	30.03
8	$\sim N(0, 1)$	$\sim N(1, 1)$	0.25	0.096	0.050	3.111	10.310	0.642	2.778	9.55
			0.50	0.494	0.392	4.339	10.881	1.373	3.027	12.72
			0.75	-0.099	0.313	3.405	11.061	1.218	1.206	15.16
			0.80	-0.079	0.383	2.527	13.058	0.715	0.413	17.86
			0.90	-0.057	0.241	2.138	17.183	0.831	0.694	21.37
			1.00	0.020	0.358	1.875	14.924	1.206	0.704	24.13
16	$\sim N(0, 0.25)$	$\sim N(1, 0.25)$	0.25	1.276	2.458	0.752	0.688	1.909	6.633	0.03
			0.50	0.544	4.229	32.236	652.725	0.623	18.482	0.10
			0.75	0.414	0.468	1.986	7.403	0.539	2.547	0.19
			0.80	0.366	0.043	1.892	5.412	0.471	0.612	0.26
			0.90	0.004	0.109	0.847	1.031	0.184	0.141	2.25
			1.00	0.005	0.047	0.847	3.441	0.183	0.115	5.61
16	$\sim N(0, 1)$	$\sim N(1, 1)$	0.25	-0.013	0.057	0.689	1.568	0.131	0.576	0.14
			0.50	-0.017	0.021	0.687	1.485	0.141	0.640	0.29
			0.75	0.020	0.167	0.861	1.085	0.131	0.276	0.49
			0.80	-0.045	0.148	0.896	1.095	0.166	0.300	0.52
			0.90	-0.049	0.134	0.890	1.081	0.171	0.245	1.33
			1.00	-0.047	0.153	0.876	0.995	0.193	0.191	2.54

in Table 4.22. The main idea behind not taking fixed true values is to see how summary measures e.g. bias, IQR and MdSE differ in comparison to measures obtained assuming true fixed values. It is found in Table 4.22 that the bias, IQR, and MdSE are unusual often as true parameters are not fixed and comes from normal distributions. The problem is severe when prior variance is comparatively larger, for example, when $n = 8, \alpha = 0.5$, true parameters are from $\beta_0 \sim N(0, 1)$ and $\beta_1 \sim N(1, 1)$, then bias, IQR and MdSE are unusual in comparison to neighbouring values. Also, it is observed in Table 4.22 that the probability of separation does not follow any increasing pattern with the increase of α while it was in increasing pattern previously in Table 4.21 when true parameters were fixed. It should be noted that as percentage of separation and summary measures were slightly less stable when number of simulations was 10,000 particularly with $\beta_0 \sim N(0, 0.25)$, and $\beta_1 \sim N(1, 0.25)$ and $n = 8$, we implemented 20,000 simulations instead of

10,000 simulations in each trial corresponding to Table 4.22.

Table 4.22: Simulated performance of estimates in Bayesian designs with 20,000 simulations

	Priors			Bias		IQR		MdSE		
n	β_0	β_1	α	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	% Sep
True parameters: $\beta_0 \sim N(0, 0.25)$, $\beta_1 \sim N(1, 0.25)$										
8	$\sim N(0, 0.25)$	$\sim N(1, 0.25)$	0.25	1.185	5.044	134.338	3072.064	3.729	28.583	8.59
			0.50	1.680	5.296	726.159	3608.230	3.078	28.045	10.01
			0.75	-0.003	4.635	5.460	21.948	1.207	21.479	26.27
			0.80	-0.003	3.631	5.503	20.701	0.303	14.094	20.36
			0.90	0.548	3.639	9.234	17.756	25.203	13.239	22.67
			1.00	-0.042	3.636	5.638	16.172	25.687	13.217	28.98
True parameters: $\beta_0 \sim N(0, 1)$, $\beta_1 \sim N(1, 1)$										
8	$\sim N(0, 1)$	$\sim N(1, 1)$	0.25	-0.096	0.053	2.525	10.034	0.641	1.629	10.78
			0.50	2.146	6.993	5.575	31.572	12.937	50.613	21.54
			0.75	-1.284	-0.710	8.344	0.925	1.649	1.200	76.44
			0.80	-1.152	-1.000	5.953	1.205	1.495	1.000	73.32
			0.90	0.637	-1.336	1.306	1.239	0.640	2.179	10.54
			1.00	0.013	0.361	1.769	14.685	1.206	0.588	24.38
True parameters: $\beta_0 \sim N(0, 0.25)$, $\beta_1 \sim N(1, 0.25)$										
16	$\sim N(0, 0.25)$	$\sim N(1, 0.25)$	0.25	0.181	2.990	7.775	647.483	1.136	25.490	0.04
			0.50	0.571	3.952	25.033	481.229	0.629	16.628	0.06
			0.75	1.099	3.828	59.620	314.848	1.207	15.041	0.27
			0.80	0.366	-0.483	0.985	0.960	0.170	0.613	0.10
			0.90	0.004	-0.273	0.802	0.648	0.175	0.141	0.70
			1.00	0.003	-0.154	0.975	0.785	0.183	0.115	2.68
True parameters: $\beta_0 \sim N(0, 1)$, $\beta_1 \sim N(1, 1)$										
16	$\sim N(0, 1)$	$\sim N(1, 1)$	0.25	0.328	1.675	1.177	8.705	0.258	2.807	1.59
			0.50	2.813	6.364	5.403	12.373	9.206	42.275	0.90
			0.75	-0.679	0.515	3.386	7.059	0.462	0.424	0.97
			0.80	2.107	-2.629	5.679	13.021	4.439	6.911	5.75
			0.90	-0.666	-0.193	0.799	0.808	0.482	0.245	1.33
			1.00	0.322	0.319	0.904	1.097	0.296	0.228	3.46

In realistic scenario practitioners might face difficulties to specify prior distribution(s) for unknown parameter(s). This limits the applicability of the Bayesian designs. Therefore, as an alternative to Bayesian designs and to cover uncertainties in the parameter(s) specifications another robust approach that can be taken into account is maximin optimality. In maximin optimality, the designs which maximize the minimum of a real valued (standardized) function of the Fisher information matrix over a range of parameter values are sought [Dette, 1997; Dette et al., 2003; Müller, 1995]. The aim of maximin optimality is to protect the design of experiment against the worst possible case [Biedermann et al., 2004]. The use of

maximin approach increases the efficiency of a design [Imhof, 2001]. A few examples of maximin designs can be found in linear models [Müller, 1995], in exponential growth and heteroscadastic polynomial models [Imhof, 2001], in regression models [Dette et al., 2003], in compartmental model [Biedermann et al., 2004] and in weighted polynomial regression models [Biedermann et al., 2004]. The applicability of maximin approach on the DP_s-optimality will be investigated in future.

4.14 Conclusions

Separation causes non-existence of maximum likelihood estimates during binary logit analysis, particularly in small samples. In this chapter, we have proposed new probability-based optimality criteria that will help minimizing the probability of separation by devising values of a design variable and thereby increasing the possibility of MLEs existence in the binary logit analysis. Further, Bayesian version of the probability based compound optimality criterion brings added benefit by having less bias, MdSE, and reduced probability of separations in comparison to non-Bayesian designs. The simulation studies verifies the ability of probability based Bayesian and non-Bayesian designs by testing the parameters of logit models.

D-optimality criterion is useful to obtain design values which enable precise parameter estimation. However, D-criterion cannot handle the issue of separation appropriately when there are small number of sample points. We have proposed new probability-based P_s-criterion that works well to minimize probability of separation. However, the relation between D- and P_s-criterion is inverse i.e. to emphasize separation issue experimenters need to compromise with the precision of the parameter estimation. Therefore, we propose a compound criterion DP_s- that will balance between precision and the separation hazard.

We have derived a theorem that explains probability of separation is less when all design values are distinct, implies that to minimize probability of separation the consideration of complete separation is enough rather than considering quasi-complete separation separately. This reduces our optimization task during the generation of design values by minimizing probability-based optimality criteria.

We have done local optimization in this study. D-optimal designs with a small sample (say $n = 4$) are not good as they may show strong separation problem. However, adequate sample size (say $n = 18$) will produce low probability of separation (e.g. separation probability lower than 0.05) (see P_s-criterion in Table 4.9). In P_s-optimal designs the probability of separation is 6.25% with sample size 8. The probability of separation reduces to 1% with 12 sample points (see P_s-criterion in Table 4.10). However, P_s-criterion cares less about precision of parameter estimates. The compound DP_s-criterion takes into account precision and separation probability simultaneously. In DP_s-optimal designs the probability of separation is less than 2.7% with 10 sample points (see Table 4.11).

Increasing number of sample points reduces the probability of separation for any design. This reduction is faster in P_s-optimal designs. If an experimenter has adequate resources he/she might think to choose D-optimal design with large number of sample points, otherwise DP_s-optimal would be the good choice for any experiment where the concept of separation gets greater priority.

D-efficiency is quite low in P_s-optimal designs and conversely P-efficiency is lower in D-optimal designs. D- and P-efficiency lines of DP_s-optimal designs passes in the middle of efficiency lines corresponding to D- and P_s-optimal designs (see Figures 4.7 and 4.8). The higher value of mixing constant in DP_s-optimal designs provides better designs that will enable good precision of parameter estimates as well as less separation.

DP_s-optimal designs with size $n \geq 16$ and mixing constant in the range $0.75 \leq \alpha < 1$ might be the good choice for any experiment as these will enable low probability of separation as well as moderate D-criterion i.e. moderately good precision (see Table 4.12). However, we have looked for optimal designs using odds of the P_s-criterion that did not provide us designs better than the designs obtained so far.

We did sensitivity analysis of the DP_s-criterion to understand the effect of changes in parameters to the magnitude of the criterion. The changes in β_1 parameter have little effect on the magnitude of DP_s-criterion when mixing constant is close to 0. On the other hand the D-optimal designs (i.e. when mixing constant is 0) are less affected by minor changes in intercept parameter β_0 (see Tables 4.13 and 4.14).

Simulation studies have been carried out to enable comparison of DP_s- designs for different values of α, different sample sizes and different values of model parameters. The robustness of DP_s- designs has been investigated to see the impact of parameter misspecifications. The mean based summary estimates e.g. bias, standard error (SE) and mean squared error (MSE) are very high because they are affected by extreme values that were originated as consequence of complete or quasi-complete separation in the simulated responses. Therefore, mean based estimates are not suitable for assessing the simulated performance of parameter estimates.

Median of estimates in the simulation study indicate that findings support the results that were conjectured in theory of separation presented in earlier sections, particularly probability of separation decreases with the increase of design size and with the decrease of mixing constant α. That means separation increases with the increase of the magnitude of α i.e. the more towards D-optimality, the more severity of separation problem. (see Table 4.15 - Table 4.19)

Locally optimal designs can be affected by parameter misspecifications. An alternative way to make the designs more robust is to assume sufficient knowledge about model parameters and thus obtain (pseudo) Bayesian designs by using Bayesian criterion. Therefore, a Bayesian version of DP_s-criterion has been developed and thereby, Bayesian designs have been obtained. It is observed that the values in the Bayesian designs are wider when priors have larger variability with respect to their magnitude of location parameters and less wider than the values of non-Bayesian designs when priors have smaller variability with respect to location parameter.

Simulation studies reveal that Bayesian designs perform better than non-Bayesian designs, particularly by providing less bias, interquartile range (IQR), median squared error (MdSE), and above all by percentage of separation problem lower than that of non-Bayesian designs. However, Bayesian designs may be impractical to practitioners wherever it becomes difficult to specify priors. As an alternative to Bayesian designs, a maximin approach can be used to obtain designs that might protect the designs from the worst possible scenario.

The newly devised probability-based DP_s-optimality criterion, either Bayesian or non-Bayesian, might be useful for a small experimental design with binary responses that might have a greater chance to have separation problem. The use of this probability based criterion might enhance the chance of existence of maximum likelihood estimates in designed experiments.

Chapter 5

Discussion and Conclusion

5.1 Introduction

As mentioned earlier, this study has dealt with two main issues: zero estimates of variance components in mixed models and separation problem that causes non-existence of maximum likelihood estimates in logit models. The zero estimates of variance components have been handled by applying Bayesian methods assuming some non-informative or weakly informative priors for relevant parameters. In addition to these, simulation studies have been carried out to assess the quality of likelihood and Bayesian estimates. Separation problem has been addressed through optimal design technique by devising probability-based non-Bayesian and pseudo Bayesian optimality criteria.

In the fuel economy experiment, researchers were unsure about the results obtained from the likelihood methods as some of the estimates of variance components were zero which were unrealistic. These led us to reanalyze fuel economy data in Bayesian methods and compare results with those from classical methods. The polypropylene industrial experiments had similar problem as in fuel economy experiments. The variance components due to batches under multi-stratum design were estimated as zero that were not logical. The earlier researchers conjectured that Bayesian analyses assuming some non-informative and weakly informative priors for the variance components would sort out the problem. During binary logit analysis of the polypropylene experiment there was convergence failure and thus non-existence of MLEs in the analysis due to the separation problem in the

data. We proposed a probability-based optimal design technique to minimize the problem of separation in experimental studies. A pseudo Bayesian version of probability based optimality criterion has also been proposed to obtain designs that is comparatively less sensitive to parameter misspecifications.

5.2 Fuel Economy Experiment

In the Bayesian analysis of fuel economy experiment, the test fuel (T) performed better than the base fuel (B) and the between day variance component estimate was 0.059 which was 0 in likelihood-based analysis. The reason for this zero variance was small number of degrees of freedom available to compute the variance component in the likelihood method. Profile likelihood and bootstrap based methods in Section 2.9 illustrate that the likelihood estimate of variance components can be zero as both of these methods have shown that lower limit of 95% confidence interval is 0 whereas the Bayesian credible interval ensures non-zero estimates of variance components (see Table 2.18). Further, Bayesian point estimates of variance component also lie impressively within the profile likelihood 95% confidence intervals. Simulation studies shows that likelihood based and Bayesian point and interval estimates perform well though average width of likelihood based 95% confidence interval can be infinity which makes variance component estimation under likelihood based methods slightly unreliable (e.g. see Table 2.20 and Table 2.21).

There might have been a benefit of switching fuel B to T than switching B1 to B2 i.e. switching from B to T impacts positively in the distance crossed by a vehicle given per gallon of fuel. Actually this is natural as fuel T performs better than fuel B. In the analysis of nested models it was revealed that there was no effect of weeks in the responses implies that ignoring the provision of gap of 4-5 days will reduce experimental duration and thereby the cost of experimentation.

The round robin experiment on fuel A shows that there is 95% guarantee that if fuel A is tested on the same laboratory they will differ by less than 10.66 and by less than 12.19 if A is tested in different laboratories (see Table 2.11). Similar conclusions are found for fuel B also. Profile likelihood and bootstrap based methods (see Table 2.19) as well as simulation studies (see Table 2.24 and Table 2.25)

reassure that Bayesian method perform well in round robin experiments.

The parameters of all models in the fuel economy and round robin experiments have passed the convergence diagnostic tests. Therefore, we may conclude that the Bayesian models are fitted well in the fuel economy experiments. The Bayesian method of analysis has enabled variance component estimation while the traditional likelihood method fails.

5.3 Polypropylene Experiment

The residual maximum likelihood (REML) and generalized least square (GLS) method gave misleading conclusions, particularly by providing estimates of variance components as zero during mixed cumulative logit analysis of data from the polypropylene experiment due to having insufficient number of whole-plots. Goos and Gilmour [2012] assumed that a Bayesian analysis considering some weakly informative priors for the variance components would be useful for estimating those inestimable variance components.

During binary logit analysis using Bayesian methods we faced severe convergence difficulties while adding some of the interactions in the models. It is not unusual to encounter convergence problems during binary or categorical data analysis as indicated by Chipman and Hamada [1996], Collett and Stepniewska [1999], Goos and Gilmour [2012]. The convergence problems led us to fit Bayesian models for binary data just by following the models simplified by Goos and Gilmour [2012]. The Bayesian estimates of mixed binary logit models are similar to those of likelihood-based estimates. For coating 3 and coating 4 the variance components due to batch are estimated as 0 whereas their Bayesian counterparts are 1.521 and 17.300.

We did independent search of Bayesian models in cumulative logit analysis of ordered categorial responses and followed stepwise manual forward selection method to select the models and compared them on the basis of deviance information criterion (DIC) values. For coating 1 the selected factors and interactions were the same both in classical and Bayesian methods. The variance component due to batch was inestimable in the classical method, however, Bayesian analysis was useful to esti-

mate the variance component which was 3.619. Similarly the variance components of coating 4 and coating 5 were 0 which were not realistic. The Bayesian methods provide these estimates as 53.760 and 2.047 respectively. The variance components measured in Bayesian methods are positive and are often more inflated than their classical counterparts. One of the reasons could be the assumed priors for variance components in Bayesian analysis. However, Gelman [2006] discussed in details how priors can affect inferences (see Section 3.6.6 and Section 3.8 for more discussion on variance components).

There were some factors and interactions that were not identified in classical methods, for example, Power×Time, EPDM×Ethylene, EPDM×Talcum for coating 2; Talcum, Lubricant, Talcum×Time, Lubricant×Power for coating 3.

The combined analysis of coatings enables us to compare factor effects across different coatings. The combined analysis reassures many of the conclusions drawn in the separate analysis of coatings and also identifies newly the main effect Lubricant and interaction effects, for instance, Ethylene×Power, C2×Ethylene, C2×EPDM, C2×Time, C4×Time found to be important. The main benefit of doing the combined analysis is to enable experimenters to examine whether the factor effects differ from coating to coating. As some of the interactions e.g. C2×EPDM, C3×Power, C3×Time, C4×Ethylene are found to be non-negligible, these imply that effects of EPDM, Power, Time and Ethylene vary from coating to coating.

Apparently there was no convergence problem during the Bayesian analysis of cumulative logit models except for coating 4. Perhaps the major reason could be the response pattern in coating 4 (see Table 3.2). Approximately three quarters of responses are 0 and the rest of them fall under response categories categories 1-5 and there is likely to be 0 cell frequencies in the contingency table regarding coating 4. Allison [2008] shows that contingency table with one or more cell frequencies 0 causes convergence problem and ultimately the non-existence of MLEs during analysis.

All selected models passed the convergence diagnostics in cumulative logit analysis of individual and combined analysis of coatings. One limitation of the Bayesian analysis of the polypropylene experiment is that we did not test the proportionality

assumptions of the coefficient of cumulative logit models though models were fitted considering the proportionality assumption. However, separate logistic fits could be compared to see the plausibility of parallelism for the data, though the test of proportionality was equally difficult during fitting models in the classical methods because of the complex nature of the models.

5.4 Optimal Design with Separation

When responses are binary, separation causes non-existence of parameter estimates in logit models specially for data from small samples. To deal with separation dilemma, the existing methods are deleting the problem variables or adjusting data and so on, but not in the light of optimal design methods. In this study we have proposed an optimal design technique to control separation at the design stage.

The newly proposed optimality criteria P_s- and DP_s- handles the separation issue well. Though P_s-criterion controls separation well at the design stage, its use would be limited as P_s- criterion loses its credibility with respect to precision of the parameter estimates. However, DP_s-optimality defined in (4.29) balances both precision and separation by considering simultaneously P_s- and D-optimality and introducing a mixing constant in the criterion. The choice of values of mixing constant (α) tends to 0 implies the emphasis is on the P_s-optimal design i.e. less separation and tends to 1 implies the emphasis is on the D-optimal design i.e. more precision of the parameter estimates.

We did local optimization in all cases using R-package *optim*. Initially we have written codes for optimizing our criteria, later found that R-package *optim* provides better design than ours with respect to separation and precision. The default method in *optim* is a derivative-free optimization routine called the Nelder-Mead simplex algorithm. However, in *optim* we have used a bound constrained option 'L-BFGS-B' that is a limited memory algorithm (see Byrd et al. [1995] for further details). We did not implement Nelder-Mead or other algorithms to choose our probability-based optimal designs as ordered design values obtained by these algorithms lead high probability of separation.

D-optimal designs presented in Table 4.9 exactly matches with the results in Atkinson et al. [2007] because we know that half of the design values are negative (-1.5434) and half of the values are positive (1.5434) as established for logistic models. The probability of separation reduces with the increasing number of sample points in D-optimal designs. In case of P_s-optimality the design values are tiny and roughly half of the values are negative and rest of them are positive (see Table 4.10). In P_s-optimal designs D-efficiency is low. Therefore, we do not recommend experimenters to use P_s-optimal designs as these will provide low precision of parameter estimates.

DP_s-optimal designs are better to handle separation and precision of parameter estimation. If researchers are satisfied to accept maximum of 5% separation then number of sample points equal to 9 or more is enough. Both P- and D-efficiencies reduces with the increasing number of sample points. Actually this happens due to comparing P-efficiency of D-optimal designs with that of P_s-optimal designs and D-efficiencies of DP_s-optimal designs with that of D-optimal designs. However, without the picture of probability of separation D- and P-efficiencies are less informative. With the increasing number of support points probability of separation reduces for all D-, P_s- and DP_s-optimal designs (see Figure 4.2).

Nobody knows the true values of model parameters before conducting an experiment. The sensitivity analysis of DP_s-criterion shows that β_0 has less effect when $\alpha \to 1$ i.e. in D-optimal designs and β_1 has less effect when $\alpha \to 0$ i.e. in P_s-optimal designs. The robustness of designs has also been assessed through simulation studies which enable comparison of DP_s-designs for different values of mixing constant, different sample sizes, and different values of model parameters. In the simulated data, percentage of separation occurs as was expected theoretically. It is evident from simulation studies that gradually complete separation increases when design moves from P_s-designs towards D-optimal designs (see Tables 4.16, 4.17, 4.18 and 4.19). Locally optimal designs can be affected by parameter misspecifications. Therefore, Bayesian version of DP_s-criterion has been developed and thereby, Bayesian designs have been obtained. Simulation studies reveal that Bayesian designs perform better than non-Bayesian designs by providing less bias, interquartile range (IQR), median squared error (MdSE), and above all by percentage of separation lower than that of non-Bayesian designs (see Tables 4.16, 4.17,

and 2.21) .

5.5 Conclusions and Future Research

The Bayesian methods implemented in this study are not to show the outright domination over classical methods. Rather Bayesian methods have been considered as complementary to classical methods where the latter failed in the analysis of data from complex designed experiments. The profile likelihood and bootstrap based confidence intervals ensures that Bayesian estimates are not absurd. Additional support comes from simulation studies which assessed the quality of point and interval estimates in Bayesian and classical methods. Therefore, the Bayesian methods that we applied in fuel efficiency experiments might be applicable in industrial experiments where there are continuous responses associated with mixed models and nested models. The analysis of polypropylene experimental data would be a good example to follow in industrial experiments involving ordered categorical responses and mixed models. Our formulas for computing probability of separation can be used before designing any study that has a binary response variable and concern for separation problem. If there are many sets of equal number design points then sequential method of computation of probability of separation can be used. DP_s-optimal designs can be used to minimize the probability of separation together with precise parameter estimation. We suggest researchers to use DP_s-optimal designs with size $n \geq 16$ and mixing constant $0.75 \leq \alpha < 1$ if there is a great concern simultaneously for separation and precision. However, if researchers have adequate resources and precise parameter estimation is the top priority, he/she can use D-optimal designs with size $n > 16$, that will lead probability of separation less than 5% (see Table 4.9). Nevertheless, if researchers spread the D-optimal points out a bit, it will not hamper the D-optimality, however, certainly will reduce the probability of separation. In addition to these, the pseudo Bayesian designs have appeared to strengthen the probability based optimal designs by making them more robust to parameter misspecifications.

We did not attempt to test the proportionality assumption of cumulative logit models. The assumption of proportionality can be tested considering separate logistic models and priors for each of the parameters. Simulation studies on ordered

logit analysis under polypropylene experiment will be accomplished in future. The MLEs existence can be compared under various traditional design criteria namely D-, A-, G-, E- and so on when responses are binary or categorical. To solve the problem of separation we have dealt with only one covariate. The research of probability-based optimal designs can be extended for multiple covariates. Also research involving separation can be extended for probit models in future, where DPs-optimal design will be different.

Appendix

The Appendix contains a sample of WinBUGS and R codes used throughout the Thesis.

WinBUGS and R Codes in Fuel Economy and Round Robin Experiments

Testing Contrast: T-B

```
model{
for (i in 1:6){
y[i]~ dnorm(mu[i],tau)
mu[i]<-base + fueldiff * fuel[i]  + v[day[i]]
}
# J is the number of days
for (j in 1:3){
v[j]~dnorm(0, prec)
}

#priors
base~dnorm(38, 0.1)
fueldiff~dnorm(0,0.001)
prec<-(1-R)*tau/R
# weakly informative essentially uniform on (0,1)
R~dbeta(1,1)
tau<-1/sig2
sig2<-sig*sig
sig<-exp(logsig)
logsig~dunif(-20,20)
sigb2<-1/prec
add<-base+fueldiff
diffpct<-100*fueldiff/base
}
```

Testing Contrast:(T-B)-(B2-B1)

```
model{
for (i in 1:12 ){
mu[i]<-f[fuel[i]]+v[day[i]]
y[i]~ dnorm(mu[i],  tau)
```

```
for (k in 1:4) {
  f[k] ~ dnorm(0,.0001)
}
# J is the number of days
for (j in 1:6) {
  v[j]~dnorm(0,prec1)
}
#priors
prec1<-(1-R)*tau/R
R~dbeta(1,1)
tau<-1/sig2
sig2<-sig*sig
sig<-exp(logsig)
sigb2<-1/prec1
logsig~dunif(-20,20)

diff1<-f[2]-f[1]
diff2<-f[4]-f[3]
difff<-diff2-diff1
}
```

Codes for Nested Model

```
model {
for (i in 1:2){    # i denotes a week
 for (j in 1: 2){  # j denotes a fuel
for (k in 1: 3) {  # k denotes a day

y[i, j, k] ~ dnorm( mu[i, j, k],  tau )
mu[i, j, k]<- mean +omega[i]+phi[i,j] + delta[i, k]
#phi[i,j]  means effect due to fuel j in week i.
# delta[i, k] is the random effect due to kth day in week i
                    }
                }
omega[i]~dnorm(0, 0.001)
phi[i, 2]~dnorm(0, 0.001)
            }
for (k in 1:3) { delta[1, k]~dnorm(0, prec1)
                        delta[2, k]~dnorm(0, prec1)  }
mean~dnorm(30, 0.001)
prec1<- (1-R)*tau/R
# highly informative with mean of 0.5
#R~dbeta(2.5,2.5)
R~dbeta(1,1)   #weakly informative essentially uniform on (0,1)
tau<-1/sig2
sig2<-sig*sig
sig<-exp(logsig)
logsig~dunif(-20,20)
sigb2<-1/prec1
#Constraints:
omega[1]<- 0
phi[1,1]<- 0       #corner-point constraint
phi[2,1]<- 0       #corner-point constraint
}

#Inits:
list(delta=structure(.Data=c(0,0,0,1,1,1), .Dim = c(2, 3)), mean=30,  logsig= -1, R=0.5)

#Scripts: Nested Model
display('log')
check('C:/Lutfor/Shell/analysis/Nested_model/model5.odc')
data('C:/Lutfor/Shell/analysis/Nested_model/datanm1.txt')
compile(1)
inits(1, 'C:/Lutfor/Shell/analysis/Nested_model/initial5.odc')
gen.inits()
```

```
update(1000)
thin.samples(15)
set(mean)
#set(mu)
set(mu[1,1,1])
set(mu[1,1,1])
set(mu[1,1,2])
set(mu[1,2,2])
 set(mu[1,2,3] )
set(mu[1,2,3] )
set(mu[2,1,1])
set(mu[2,1,1])
set(mu[2,1,2])
set(mu[2,2,2])
 set(mu[2,2,3])
set(mu[2,2,3])
set(omega)
set(tau)
set(phi)
set(sigb2)
set(delta)
set(R)
set(prec1)
dic.set()
beg(1000)
update(30000)
stats(*)
dic.stats()
density(*)
#trace(*)
autoC(*)
history(*)
```

Codes of Round Robin Analysis for Fuel A

```
model {
for (i in 1:15) {
mu[i]<-mean  + u[lab[i]]
y[i]~ dnorm(mu[i], tau)
}
# J is the number of labs
for (j in 1:10) {
u[ j ]~dnorm(0,tauL)
}
#priors
logsig~dunif(-20,20)
sig<-exp(logsig)
sig2<-sig*sig
tau<-1/sig2
mean~dnorm(-6,0.0001)
# highly informative with mean of 0.5
#R~dbeta(2.5,2.5)
# weakly informative essentially uniform on (0,1)
R~dbeta(1,1)

tauL<-(1-R)*tau/R
sigL2<-1/tauL
sigL<-sqrt(sigL2)
reprodSD<-sqrt(sigL2+sig2)
reprod<-2.8*reprodSD
repeat<-2.8*sig
}
#Inits:
list(u=c(0,0,0,0,0,0,0,0,0,0), mean=20, R=0.5, logsig=-1)
```

R Codes for Wald, Profile Likelihood and Bootstrap based Confidence Intervals for Fuel A

```
library(MASS)
library(nlme) # will be loaded automatically if omitted
library(lme4)
data2<-read.table("data_T_B.txt",header=T)
data2=data.frame(data2)
data2
outputfuel<- glmer(y ~ fuel +(1 | day), data = data2, family=gaussian,REML=FALSE)
outputfuel
summary(outputfuel)
confint(outputfuel,method="profile")
confint(outputfuel,method="Wald")
confint(outputfuel,method="boot",nsim=500,oldNames=FALSE)
```

R Codes for Simulation Studies in Likelihood based Method for Fuel A

```
NS=2000
nc1=3
CEsti=matrix(NA,nrow=NS,ncol=nc1)
CBias=matrix(NA,nrow=NS,ncol=nc1)
CBiasSq=matrix(NA,nrow=NS,ncol=nc1)
L=matrix(NA,nrow=NS,ncol=nc1)
U=matrix(NA,nrow=NS,ncol=nc1)
W=matrix(NA,nrow=NS,ncol=nc1)
CTF=matrix(NA,nrow=NS,ncol=nc1)
alpha=32
beta2=1.4
sigSq=0.05            #error variance
sigma=sqrt(sigSq)     #error sd
sigSqB=0.05           #variance component due to random effect day
sigB=sqrt(sigSqB)     #variance component due to random effect day in terms of sd
TP=c(alpha, beta2, sigSqB)  #True parameters
nday=3  #number of days
day=rep(1:nday,each=2)
#creating data corresponding to fuel
fuel<-c(rep(0,nday),rep(1,nday)) #0 means B and 1 means T
stop.iter1<-0
stop.iter2<-0
for (s in 1:NS)
{
#creating random effects due to day
#set.seed(234+30*i)
del<-rnorm(nday,0,sigB)
delta<-rep(del,each=2)
delta
#assuming effect of treatment (fuel)
beta=c(rep(0,nday),rep(beta2,nday))   # 0 if for B and beta2 is for T
mu=alpha+beta+delta
y<-NULL  #this should be kept out of the loop

for(k in 1:length(day)) {
y[k]<- rnorm(1,mu[k],sigma)  #generating single random number from normal distribution with
                             #mean mu[k] and variance 0.05
              }
data1=cbind(y,day,fuel)
data2=data.frame(data1)  #making data frame for anlaysis in R

library(nlme)
simuTB.lme2=tryCatch(lme(y~fuel, random= ~1 | day,data=data2),error=function(e) e)
if(!inherits(simuTB.lme2,"error")) {
simuTB.lme3<-summary(simuTB.lme2)
```

```
#Finding intervals inclding the estimates
hh=tryCatch(intervals(simuTB.lme3),error=function(e) e)
if(!inherits(hh,"error")) {
#Extracting fixed effecst from linear mixed model.
fixef=hh[[1]]
#Extracting random effect from linear mixed model.
ranef1=hh[[2]]$day
ranef=ranef1^2     #converting random effect estimate sd to variance
#Fixed and random effects with intervals
fixran=rbind(fixef, ranef)
fixran=as.matrix(fixran)
PEst=fixran[,2]     #parameter estimates
CEsti[s,]=PEst      #parameter estimates in ith simulation
LB=fixran[,1]       #lower bound of estimates (from 95%intervals)
UP=fixran[,3]       #upper bound of estimates (from 95%intervals)
L[s,]=LB  #lower bound of estimates in ith simulation (from 95%intervals)
U[s,]=UP  #upper bound of estimates in ith simulation (from 95%intervals)
B=PEst-TP #Bias of the estimates
CBias[s,]=B #Bias of the estimates in the ith simulations
CBiasSq[s,]=B^2 #Bias square of the estimates in the ith
                #simulations to be used in MSE calculations
W[s,]=UP-LB #width in sth simulation
D=(TP>=LB & TP<=UP)  #Determination of true false for coverage prob. calcluation
CTF[s,]=D  #True false matrix for coverage prob. calcluation in ith simulation
} else{stop.iter1<-stop.iter1+1}
} else{stop.iter2<-stop.iter2+1}
}
#CEsti
CEstimates=apply(CEsti, 2, mean,na.rm=TRUE)
CEstimates
CEstSD=apply(CEsti,2,sd,na.rm=TRUE)
#CBias
CBiasEst=apply(CBias, 2, mean,na.rm=TRUE)
CBiasEst
CRelBias=(CBiasEst/TP)*100  #Relative bias in percetage
CRelBias
#CBiasSq
CMSE=apply(CBiasSq, 2, mean,na.rm=TRUE)
CMSE
CRMSE=sqrt(CMSE)
CRMSE
CL=apply(L, 2, mean,na.rm=TRUE)  #mean of lower bounds in 95% CI
CU=apply(U, 2, mean,na.rm=TRUE)  #mean of upper bounds in 95% CI
#***** Median based estimates **************
CMdEst=apply(CEsti, 2, median,na.rm=TRUE)
CMdBias=apply(CBias, 2, median,na.rm=TRUE)
CMdSE=apply(CBiasSq, 2, median,na.rm=TRUE)
CRMdSE=sqrt(CMdSE)
CMdRelBias=(CMdBias/TP)*100
CMeanWidth=apply(W, 2, mean,na.rm=TRUE)  #mean of width of 95% CI
CMdWidth=apply(W, 2, median,na.rm=TRUE)  #median of width of 95% CI
CCP=apply(CTF, 2, mean,na.rm=TRUE)  #coverage prob.
CMinEst=apply(CEsti, 2, min,na.rm=TRUE)
CMaxEst=apply(CEsti, 2, max,na.rm=TRUE)
CWidthEst=CMaxEst-CMinEst
library(xtable)
Cblank=rep(0,times=length(CMdEst))
CEstMatrix=cbind(CBiasEst,CRelBias,CRMSE,Cblank,CMdBias,CMdRelBias,CRMdSE,CCP,
CMeanWidth,CMdWidth)
library(xtable)
xtable(CEstMatrix,digit=c(1,3,2,3, 1,  3,2,3, 3,3,3))
stopping_rate1=stop.iter1/NS
stopping_rate1
stopping_rate2=stop.iter2/NS
stopping_rate2
```

R Codes for Simulation Studies in Bayesian Method for Fuel A

```
NS=2000
nc2=3
#***Bayesian initial matrices
BEst=matrix(0,nrow=NS,ncol=nc2)
BBias=matrix(0,nrow=NS,ncol=nc2)
BBiasSq=matrix(0,nrow=NS,ncol=nc2)
BTF=matrix(0,nrow=NS,ncol=nc2)
BEst2.5P=matrix(0,nrow=NS,ncol=nc2)
BEst97.5P=matrix(0,nrow=NS,ncol=nc2)
BMedEst=matrix(0,nrow=NS,ncol=nc2)
BMedBias=matrix(0,nrow=NS,ncol=nc2)
BMedBiasSq=matrix(0,nrow=NS,ncol=nc2)
varcomp=0.05
varcompsd=sqrt(varcomp)
for (s in 1:NS)
{
day=rep(1:3,each=2)
#creating data corresponding to fuel
fuel<-c(rep(0,3),rep(1,3)) #0 means B and 1 means T
#creating random effects due to day
#set.seed(234+30*i)
del<-rnorm(3,0,varcompsd)
delta<-rep(del,each=2)
#assuming effect of treatment (fuel)
#fuel effect due to B is 0 and effect due to T is 1.5
beta=c(rep(0,3),rep(1.4,3))   # 0 if for B and 1.5 is for T
#***estimated reponse/generating normal response
alpha=32
mu=alpha+beta+delta
y<-NULL  #this should be kept out of the loop
for(k in 1:length(day)) {
y[k]<- rnorm(1,mu[k],varcompsd)  #generating single random number from
                                  normal distribution with
                                 #mean mu[k] and variance 0.05
             }
data1=cbind(y,day,fuel)
data2=data.frame(data1)  #making data frame for anlaysis in R

#***********************Bayesian programme*******************
library(R2WinBUGS)
# Using the following cat("  ....model here....", file=".......")
#command we can include model in the R directory.
data2=data.frame(data1)  #making data frame for anlaysis in R
y <- data2$y
day <- data2$day
fuel<- data2$fuel
data <- list ("y" , "day", "fuel")
cat("
model{
for (i in 1:6)
{
y[i]~ dnorm(mu[i],tau)
mu[i]<-base + fueldiff * fuel[i]  + v[day[i]]
}
# J is the number of days
for (j in 1:3) {
v[j]~dnorm(0, prec)
}
#priors
base~dnorm(38, 0.1)
fueldiff~dnorm(0,0.001)
prec<-(1-R)*tau/R
# weakly informative essentially uniform on (0,1)
```

```
R~dbeta(1,1)
tau<-1/sig2
sig2<-sig*sig
sig<-exp(logsig)
logsig~dunif(-20,20)
vcomday<- 1/prec
#vcSdBayes<-sqrt(vcomday)
}", file="ex5.bug")
###initial values for the chain:
inits <- list(base=38, fueldiff=0.1, R=0.5, logsig=-1)
#parameters = c("base", "fueldiff", "tau", "prec1", "vcomday", "vcSdBayes")
parameters = c("base", "fueldiff", "vcomday")
### To start the MCMC we have to write the following commands:
fuel.sim2 <- bugs(data, inits, parameters, model="ex5.bug",
n.chains = 4, n.iter = 5000, n.burnin=2000, n.thin=15,
bugs.directory = "C:/WinBUGS14/")
#See further commands to obtain some required output.
BO<- print(fuel.sim2)  #Bayesian output (BO)
#to separte estimates from the above
BOMatrix1=BO[10]$summary
BOMatrix2=as.matrix(BOMatrix1)
BOMatrix=BOMatrix2[-4,]
BEstim1=BOMatrix[,1]
BEst[s,]=BEstim1
BE2.5P=BOMatrix[,3]
BEst2.5P[s,]=BE2.5P
BE97.5P=BOMatrix[,7]
BEst97.5P[s,]=BE97.5P
TP=c(alpha,1.4,varcomp)
TF1=(TP>=BE2.5P & TP<=BE97.5P)
BTF[s,]=TF1
BB=BEstim1-TP
BBSq1=BB^2
BBias[s,]=BB
BBiasSq[s,]=BBSq1
BMedEst1=BOMatrix[,5]
BMedEst[s,]=BMedEst1
BBMed=BMedEst1-TP
BBMedSq1=BBMed^2
BMedBias[s,]=BBMed
BMedBiasSq[s,]=BBMedSq1
}
BEstimates=apply(BEst, 2, mean)
BBiasEst=apply(BBias, 2, mean)
RelBBias=(BBiasEst/TP)*100  #Relative Bayes bias in percetage
BMSE=apply(BBiasSq, 2, mean)
BRMSE=sqrt(BMSE)
BCP=apply(BTF, 2, mean) #Bayesian est coverage prob.
BMinEst=apply(BEst, 2, min)
BMaxEst=apply(BEst, 2, max)
BWidthEst=BMaxEst-BMinEst
BWidthEst
BwidthCI=BEst97.5P-BEst2.5P
BAvewidthCI=apply(BwidthCI, 2, mean)
BMedwidthCI=apply(BwidthCI, 2, median)
#*******Median  based estiamtes********
BMedEstimates=apply(BMedEst, 2, mean)
BMedBiasEst=apply(BMedBias, 2, mean)
RelBMedBias=(BMedBiasEst/TP)*100  #Relative Bayes Median bias in percetage
BMdMSE=apply(BMedBiasSq, 2, mean)
BMdRMSE=sqrt(BMdMSE)
#*******Final estiamtes in matrix ********
BEstMatrix1=cbind(BEstimates,BBiasEst,RelBBias,BMSE, BRMSE,BCP,BAvewidthCI,
BWidthEst)
BEstMatrix2=cbind(BBiasEst,RelBBias,BRMSE,BCP,BAvewidthCI)
BEstMatrix3=cbind(BMedEstimates,BMedBiasEst,RelBMedBias,BMdMSE, BMdRMSE,BCP,
BAvewidthCI, BWidthEst)
```

```
BEstMatrix4=cbind(BMedBiasEst,RelBMedBias, BMdRMSE,BCP,BAvewidthCI)
library(xtable)
Bblank=rep(0,times=length(BMedBiasEst))
BEstMatrix5=cbind(BBiasEst,RelBBias,BRMSE, Bblank, BMedBiasEst,RelBMedBias,
BMdRMSE, BCP,BAvewidthCI,BMedwidthCI)
xtable(BEstMatrix5,digit=c(1,3,2,3, 1,  3,2,3, 3,3,3))
```

R Codes for Simulation Studies in Bayesian Method for Round Robin Experiments

```
NS=2000
nc1=5
#***Bayesian initial matrices
BMeanEst=matrix(0,nrow=NS,ncol=nc1)
BiasEst=matrix(0,nrow=NS,ncol=nc1)
BiasSq=matrix(0,nrow=NS,ncol=nc1)
BMedEst=matrix(0,nrow=NS,ncol=nc1)
BSd=matrix(0,nrow=NS,ncol=nc1)
BMedBias=matrix(0,nrow=NS,ncol=nc1)
BMedBiasSq=matrix(0,nrow=NS,ncol=nc1)
BEst2.5P=matrix(0,nrow=NS,ncol=nc1)
BEst97.5P=matrix(0,nrow=NS,ncol=nc1)
BTF=matrix(0,nrow=NS,ncol=nc1)
BRhat=matrix(0,nrow=NS,ncol=nc1)
for (s in 1:NS){
lb=20  #number of labs
alpha=22
sigmaSq=16
sigma=sqrt(sigmaSq)
sigmaSqL=5
sigmaL=sqrt(sigmaSqL)
reprodSD<-sqrt(sigmaSq+sigmaSqL)
reprod<-2.8*reprodSD
repeat1<-2.8*sigma
TP=c(alpha,sigmaSq,sigmaSqL,repeat1,reprod)  #true parameters
lab=rep(1:lb,each=2) #number of observations in labs
#creating random effects due to day
#set.seed(234)
rel<-rnorm(lb,0,sigmaL)   #lb random effects for lb number of labs
relwrep<-rep(rel,each=2)   #lb random effects for lb number of labs with 2 repeats
#***estimated reponse/generating normal response
rows=length(relwrep)
mu=alpha+relwrep
y<-NULL  #this should be kept out of the loop
for(k in 1:rows) {
y[k]<- rnorm(1,mu[k],sigma)  #generating single random number from
                             normal distribution with
                             #mean mu[k] and sd 0.05
             }
data1=cbind(y,lab)
library(R2WinBUGS)
data2=data.frame(data1)  #making data frame for anlaysis in R
y <- data2$y
lab <- data2$lab
data <- list ("y", "lab")
cat("
model { for (i in 1:40)
{
mu[i]<-mean + u[lab[i]]
y[i]~ dnorm(mu[i], tau)}
# J is the number of labs
for (j in 1:20) {
u[ j ]~dnorm(0,tauL)}
#priors
logsig~dunif(-20,20)
```

```
sig<-exp(logsig)
sig2<-sig*sig
tau<-1/sig2
mean~dnorm(20,0.001)
R~dbeta(1.5,1.5)
tauL<-(1-R)*tau/R
sigL2<-1/tauL
sigL<-sqrt(sigL2)
reprodSD<-sqrt(sigL2+sig2)
reprod<-2.8*reprodSD
repeat<-2.8*sig
}", file="exRR.bug")
###initial values for the chain:
inits <- list(u=rep(0,lb), mean=20, R=0.5, logsig=-1)
parameters = c("mean", "sig2", "sigL2", "repeat", "reprod")
### To start the MCMC we have to write the following commands:
fuel.simRR <- bugs(data, inits, parameters, model="exRR.bug",
 n.chains = 4, n.iter = 5000, n.burnin=2000, n.thin=10, bugs=TRUE,
 bugs.directory = "C:/WinBUGS14/")
#See further commands to obtain some required output.
RRB<- print(fuel.simRR)  #Bayesian output of Round Robin (RRB)
#to separte individual estimates from the above
RRBM1=RRB$summary  #summary in matrix form
RRBM2=RRBM1[-6,]    #deleting deviance
#RRBM3=RRBM1[c(-4,-5,-6),]  #keeping the parameters that will be
assessed for bias and others
#Round robin robustness assessment:
RRBM=as.matrix(RRBM2)
BMeanEst1=RRBM[,1]
BMeanEst[s,]=BMeanEst1
BSdev=RRBM[,2]
BSd[s,]=BSdev
BE2.5P=RRBM[,3]
BEst2.5P[s,]=BE2.5P
BE97.5P=RRBM[,7]
BEst97.5P[s,]=BE97.5P
BRhat1=RRBM[,8]
BRhat[s,]=BRhat1
BB=BMeanEst1-TP
BBSq=BB^2
BiasEst[s,]=BB
BiasSq[s,]=BBSq
#median based
BMdEstim1=RRBM[,5]
BMedEst[s,]=BMdEstim1
BMdBias=BMdEstim1-TP
BMdBiasSq1=BMdBias^2
BMedBias[s,]=BMdBias
BMedBiasSq[s,]=BMdBiasSq1
TF1=(TP>=BE2.5P & TP<=BE97.5P)
BTF[s,]=TF1
}
#mean based estiamtes:
BMeanEst
BMeanEstimates=apply(BMeanEst,2,mean)
BSd
BSdEst=apply(BSd,2,mean)
BiasEst
BBiasEst=apply(BiasEst,2,mean)
RelBias=(BBiasEst/TP)*100
BiasSq
BMSE=apply(BiasSq,2,mean)
BRMSE=sqrt(BMSE)
#median based estimates
RRBMedEstimates=apply(BMedEst, 2, mean)  #in futuer median instead of
                                         mean can be considered
BMedBias
```

```
BMedBiasEst=apply(BMedBias, 2, mean)
RelBMedBias=(BMedBiasEst/TP)*100  #Relative Bayes bias in percetage
BMedSE=apply(BMedBiasSq, 2, mean)
BRMedSE=sqrt(BMedSE)
BCP=apply(BTF, 2, mean) #Bayesian est coverage prob.
BWidth95CI=BEst97.5P-BEst2.5P
BAveWidth95CI=apply(BWidth95CI, 2, mean)
BMedwidthCI=apply(BWidth95CI, 2, median)
BMinMeanEst=apply(BMeanEst, 2, min)
BMaxMeanEst=apply(BMeanEst, 2, max)
BWidthMeanEst=BMaxMeanEst-BMinMeanEst
BWidthMeanEst  #range of the mean estimates
BMinMdEst=apply(BMedEst, 2, min)
BMaxMdEst=apply(BMedEst, 2, max)
BWidthMdEst=BMaxMdEst-BMinMdEst
BWidthMdEst  #range of the median estimates
BRhatEst=apply(BRhat, 2, mean)
RRBMeanEstMatrix1=cbind(BMeanEstimates,BSdEst,BBiasEst,RelBias, BMSE,
BRMSE,BCP,BAveWidth95CI, BWidthMeanEst,BRhatEst)
RRBMeanEstMatrix2=cbind(BBiasEst,RelBias, BRMSE,BCP,BAveWidth95CI,BRhatEst)
RRBMdEstMatrix3=cbind(RRBMedEstimates,BMedBiasEst,RelBMedBias,BMedSE, BRMedSE,
BCP,BMedwidthCI, BWidthMdEst,BRhatEst)
RRBMdEstMatrix4=cbind(BMedBiasEst,RelBMedBias, BRMedSE,BCP,BAveWidth95CI, BRhatEst)
library(xtable)
xtable(RRBMeanEstMatrix1,digit = c(1,3,3,3, 3,3,3, 3,3,3))
Bblank=rep(0,times=length(BMedBiasEst))
BEstMatrix5=cbind(BBiasEst,RelBias, BRMSE, Bblank, BMedBiasEst,RelBMedBias,
BRMedSE, BCP,BAveWidth95CI,BMedwidthCI)
xtable(BEstMatrix5,digit=c(1,3,2,3, 1,  3,2,3, 3,3,3))
```

WinBUGS Codes for Mixed Binary Logit Analysis of Coating 2

.

```
model {
    for(i in 1:300) {
      y[i] ~dbern(p[i])
     logit(p[i]) <- b0 + b1*epdm[i]  + b2*ethln[i]  + b3*talcum[i]+b4*mica[i]
                  + b5 * power[i]+ b6 * time[i]+ b7*gastype[i]+ b8*actgas[i]
                  +  b9 * power[i]*actgas[i] + u[run[i]] + v[batch[i]]
                    }
            # M groups of runs
              for (j in 1:100) {
                 u[j] ~dnorm(0, tau1)
                                }
            # L is the number of batches
               for (k in 1:20) {
                 v[k] ~dnorm(0, tau2)
                               }
               b0 ~ dnorm(0, .01)
               b1 ~ dnorm(0, .01)
               b2 ~ dnorm(0, .01)
               b3 ~ dnorm(0, .01)
               b4 ~ dnorm(0, .01)
               b5 ~ dnorm(0, .01)
               b6 ~ dnorm(0, .01)
               b7 ~ dnorm(0, .01)
               b8 ~ dnorm(0, .01)
               b9 ~ dnorm(0, .01)
            tau1~dgamma(1,1)
            rho~dbeta(5, 5)
            tau2<- (1-rho)* (tau1)/(rho)
            sigsq1 <- 1 / (tau1)
```

```
        sigsq2 <- 1 /(tau2)
    }
```

WinBUGS Codes for Mixed Cumulative Logit Analysis of Coating 2

```
model {
for (i in 1:300) {
for (k in 1:6) {
y[i, k]<- equals(astm[i], k) } }
for (i in 1:300) {
y[i, 1:6] ~dmulti(p[i, 1:6], 1)
p[i, 1]<-  1- q[i, 1]
p[i, 2]<-  q[i, 1] - q[i, 2]
p[i, 3]<-  q[i, 2] - q[i, 3]
p[i, 4]<-  q[i, 3] - q[i, 4]
p[i, 5]<-  q[i, 4] - q[i, 5]
p[i, 6]<-  q[i, 5]
logit(q[i,1])<- (a1 +    b[1] * epdm[i] + b[2] * ethln[i] + b[3] * talcum[i] + b[4] * mica[i]
                + b[5]* power[i] + b[6]* time[i] + b[7] * gastype[i] + b[8] * actgas[i]
                + b[9]* epdm[i]*ethln[i]+ b[10]*epdm[i]*talcum[i] + b[11]*power[i]*gastype[i]
                + b[12]*power[i]*actgas[i]+ b[13]*power[i]*time[i] + b[14]* ethln[i]*power[i]+
                b[15]* mica[i]*actgas[i] + u[run[i]] + v[batch[i]]  )

logit(q[i,2])<- (a2  +   b[1] * epdm[i] + b[2] * ethln[i] + b[3] * talcum[i] + b[4] * mica[i]
                + b[5]* power[i] + b[6]* time[i] + b[7] * gastype[i] + b[8] * actgas[i]
                + b[9]* epdm[i]*ethln[i]+ b[10]*epdm[i]*talcum[i] + b[11]*power[i]*gastype[i]
                + b[12]*power[i]*actgas[i]+ b[13]*power[i]*time[i] + b[14]* ethln[i]*power[i]
                +  b[15]* mica[i]*actgas[i] + u[run[i]] + v[batch[i]] )

logit(q[i,3])<-  (a3  +   b[1] * epdm[i] + b[2] * ethln[i] + b[3] * talcum[i] + b[4] * mica[i]
                + b[5]* power[i] + b[6]* time[i] + b[7] * gastype[i] + b[8] * actgas[i]
                + b[9]* epdm[i]*ethln[i]+ b[10]*epdm[i]*talcum[i] + b[11]*power[i]*gastype[i]
                + b[12]*power[i]*actgas[i]+ b[13]*power[i]*time[i] + b[14]* ethln[i]*power[i]
                + b[15]* mica[i]*actgas[i] + u[run[i]] + v[batch[i]]  )

logit(q[i,4])<-  (a4  +    b[1] * epdm[i] + b[2] * ethln[i] + b[3] * talcum[i] + b[4] * mica[i]
                + b[5]* power[i] + b[6]* time[i] + b[7] * gastype[i] + b[8] * actgas[i]
                + b[9]* epdm[i]*ethln[i]+ b[10]*epdm[i]*talcum[i] + b[11]*power[i]*gastype[i]
                + b[12]*power[i]*actgas[i]+ b[13]*power[i]*time[i] + b[14]* ethln[i]*power[i]
                +  b[15]* mica[i]*actgas[i] + u[run[i]] + v[batch[i]] )
logit(q[i,5])<-  (a5 +   b[1] * epdm[i] + b[2] * ethln[i] + b[3] * talcum[i] + b[4] * mica[i]
                + b[5]* power[i] + b[6]* time[i] + b[7] * gastype[i] + b[8] * actgas[i]
                + b[9]* epdm[i]*ethln[i]+ b[10]*epdm[i]*talcum[i] + b[11]*power[i]*gastype[i]
                + b[12]*power[i]*actgas[i]+ b[13]*power[i]*time[i] + b[14]* ethln[i]*power[i]
                +  b[15]* mica[i]*actgas[i] + u[run[i]] + v[batch[i]] )
                                }

        # M groups of runs
            for (j in 1:100) {
                                u[j] ~dnorm(0, tau1)
            }
        # L is the number of batches
                        for (k in 1:20) {
                                v[k] ~dnorm(0, tau2)
             }
            a1 ~ dnorm(0, 1.0E-04)I(a2, )
            a2 ~ dnorm(0, 1.0E-04)I(a3, a1)
            a3 ~ dnorm(0, 1.0E-04)I(a4, a2)
            a4 ~ dnorm(0, 1.0E-04)I(a5, a3)
            a5 ~ dnorm(0, 1.0E-04)I( , a4)
                       for(m in 1:15) {
            b[m] ~ dnorm(0, 1.0E-04)
            }
```

```
            tau1~dgamma(1,1)
            rho~dbeta(5, 5)
            tau2<- (1-rho)* (tau1)/(rho)
            sigsq1 <- 1 / (tau1)
            sigsq2 <- 1 /(tau2)
                    }
#Initial values:
list(a1=0.05, a2 = 0.04, a3 = 0.03, a4 = 0.02, a5 = 0, b=c(0,0,0,  0,0,0,
     0,0,0,  0,0,0,   0,0,0))
```

R Codes to Compute Complete Separation

```
function (x1, x2, x3, x4, x5, x6)
{
    y <- c(0, 0, 0, 1, 1, 1)
    x <- c(x1, x2, x3, x4, x5, x6)
    yx <- cbind(y, x)
    pf1 <- 1/(1 + exp(x1))
    ps1 <- 1 - pf1
    pf2 <- 1/(1 + exp(x2))
    ps2 <- 1 - pf2
    pf3 <- 1/(1 + exp(x3))
    ps3 <- 1 - pf3
    pf4 <- 1/(1 + exp(x4))
    ps4 <- 1 - pf4
    pf5 <- 1/(1 + exp(x5))
    ps5 <- 1 - pf5
    pf6 <- 1/(1 + exp(x6))
    ps6 <- 1 - pf6
    PCSP <- (pf1 * pf2 * pf3 * ps4 * ps5 * ps6 + ps1 * ps2 *
              ps3 * pf4 * pf5 * pf6)
    pf <- c(pf1, pf2, pf3, pf4, pf5, pf6)
    ps <- c(ps1, ps2, ps3, ps4, ps5, ps6)
    print(yx)
    print(pf)
    print(ps)
    print(PCSP)
}
```

R Codes to Compute Quasi-complete Separation

```
function (x1, x2, x3, x4, x5, x6)
{
    y <- c(0, 0, 0, 1, 1, 1)
    x <- c(x1, x2, x3, x4, x5, x6)
    yx <- cbind(y, x)
    pf1 <- 1/(1 + exp(x1))
    ps1 <- 1 - pf1
    pf2 <- 1/(1 + exp(x2))
    ps2 <- 1 - pf2
    pf3 <- 1/(1 + exp(x3))
    ps3 <- 1 - pf3
    pf4 <- 1/(1 + exp(x4))
    ps4 <- 1 - pf4
    pf5 <- 1/(1 + exp(x5))
    ps5 <- 1 - pf5
    pf6 <- 1/(1 + exp(x6))
    ps6 <- 1 - pf6
    PCSP <- (pf1 * pf2 * pf3 * ps4 * ps5 * ps6 + pf1 * pf2 *
        ps3 * pf4 * ps5 * ps6 + ps1 * ps2 * pf3 * ps4 * pf5 *
        pf6 + ps1 * ps2 * ps3 * pf4 * pf5 * pf6)
    pf <- c(pf1, pf2, pf3, pf4, pf5, pf6)
```

```
    ps <- c(ps1, ps2, ps3, ps4, ps5, ps6)
    print(yx)
    print(pf)
    print(ps)
    print(PCSP)
}
```

R Codes for Computation of DP_s-optimality when mixing constant α=0.50

```
> psdc.alpha
function(g, alpha){
##Programme # 70 #####
s <- NULL; f<- NULL; x<-NULL; d<-NULL; w<-NULL; pcs<- NULL
n<- length(g)
#alpha<- 0.75
x[1] <- g[1]
for(k in 1:(n-1))  {
                x[k+1]<- x[k]+g[k+1]
                  }
        for(i in 1:n)  {
                s[i]<- 1/(1+exp(-x[i]))
                f[i]<- 1- s[i]  }
#to see the results:
#print(s)
#print(f)
#Cumulative product of s
cps<-cumprod(s)
#Cumulative product of f
cpf<-cumprod(f)
#To see the results:
#print(cps)
#print(cpf)
for(k in 1:(n-1))  {
                pcs[k]<- (cpf[k]*(cps[n]/cps[k]) + cps[k]*(cpf[n]/cpf[k]))
                   }
#to see the results:
#print(pcs)
#prob of separation:
ps<- cps[n]+sum(pcs)+cpf[n]
#to see the results:
#print(ps)
#****************************
##Finding D-optimal design with ordered x values.
#to see the results:
#print(s)
#print(f)
x11<-rep(1, n)
x12<-x
X<-cbind(x11,x12)
for(t in 1:n){
            w[t]<-(1/n)
              }
wp<-w*s*f
W<-diag(wp)
Xt<-t(X)
INF<-Xt %*% W %*% X
DINF<-det(INF)
DCO<-(1/DINF)^(alpha/2)  # this 2 is for no. of parameters
PSO<-(ps)^(1-alpha)
#Combined criterion
DCPS<-(PSO)*(DCO)
DCPS
}
```

```
#*******************************
#When alpha=0.5
#*******************************
#**Example1
t<-optim(c(-0.1,0.001), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 1)), alpha=0.5)
t1<-rv(t$par)
ps1<-ps(t1)
dc1<-DC1(t1)
r1<-c(t1, rep(0,(20-length(t1))))   #20 is the highest length of largest design
#**Example2
t<-optim(c(-0.1,rep(0.001, 2)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 2)), alpha=0.5)
t2<-rv(t$par)
ps2<-ps(t2)
dc2<-DC1(t2)
r2<-c(t2, rep(0,(20-length(t2))))

#**Example3
t<-optim(c(-0.1,rep(0.001, 3)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 3)), alpha=0.5)
t3<-rv(t$par)
ps3<-ps(t3)
dc3<-DC1(t3)
r3<-c(t3, rep(0,(20-length(t3))))

#**Example4
t<-optim(c(-0.1,rep(0.001, 4)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 4)), alpha=0.5)
t4<-rv(t$par)
ps4<-ps(t4)
dc4<-DC1(t4)
r4<-c(t4, rep(0,(20-length(t4))))
#**Example5
t<-optim(c(-0.1,rep(0.001, 5)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 5)), alpha=0.5)
t5<-rv(t$par)
ps5<-ps(t5)
dc5<-DC1(t5)
r5<-c(t5, rep(0,(20-length(t5))))

#**Example6
t<-optim(c(-0.1,rep(0.001, 6)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 6)), alpha=0.5)
t6<-rv(t$par)
ps6<-ps(t6)
dc6<-DC1(t6)
r6<-c(t6, rep(0,(20-length(t6))))

#**Example7
t<-optim(c(-0.1,rep(0.001, 7)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 7)), alpha=0.5)
t7<-rv(t$par)
ps7<-ps(t7)
dc7<-DC1(t7)
r7<-c(t7, rep(0,(20-length(t7))))

#**Example8
t<-optim(c(-0.1,rep(0.001, 8)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 8)), alpha=0.5)
t8<-rv(t$par)
ps8<-ps(t8)
dc8<-DC1(t8)
r8<-c(t8, rep(0,(20-length(t8))))

#**Example9
t<-optim(c(-0.1, rep(0.001, 9)), psdc.alpha, NULL, method="L-BFGS-B",
```

```
lower=c(-20,rep(0.0001, 9)), alpha=0.5)
t9<-rv(t$par)
ps9<-ps(t9)
dc9<-DC1(t9)
r9<-c(t9, rep(0,(20-length(t9))))

#**Example10
t<-optim(c(-0.1,rep(0.001, 10)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 10)), alpha=0.5)
t10<-rv(t$par)
ps10<-ps(t10)
dc10<-DC1(t10)
r10<-c(t10, rep(0,(20-length(t10))))

#**Example11
t<-optim(c(-0.1,rep(0.001, 11)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 11)), alpha=0.5)
t11<-rv(t$par)
ps11<-ps(t11)
dc11<-DC1(t11)
r11<-c(t11, rep(0,(20-length(t11))))

#**Example12
t<-optim(c(-0.1,rep(0.001, 12)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 12)), alpha=0.5)
t12<-rv(t$par)
ps12<-ps(t12)
dc12<-DC1(t12)
r12<-c(t12, rep(0,(20-length(t12))))

#**Example13
t<-optim(c(-0.1,rep(0.001, 13)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 13)), alpha=0.5)
t13<-rv(t$par)
ps13<-ps(t13)
dc13<-DC1(t13)
r13<-c(t13, rep(0,(20-length(t13))))

#**Example14
t<-optim(c(-0.1, rep(0.001, 14)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 14)), alpha=0.5)
t14<-rv(t$par)
ps14<-ps(t14)
dc14<-DC1(t14)
r14<-c(t14, rep(0,(20-length(t14))))

#**Example15
t<-optim(c(-0.1, rep(0.001, 15)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 15)), alpha=0.5)
t15<-rv(t$par)
ps15<-ps(t15)
dc15<-DC1(t15)
r15<-c(t15, rep(0,(20-length(t15))))

#**Example 16
t<-optim(c(-0.1, rep(0.001, 16)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 16)), alpha=0.5)
t16<-rv(t$par)
ps16<-ps(t16)
dc16<-DC1(t16)
r16<-c(t16, rep(0,(20-length(t16))))

#**Example 17
t<-optim(c(-0.1, rep(0.001, 17)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 17)), alpha=0.5)
t17<-rv(t$par)
ps17<-ps(t17)
```

```
dc17<-DC1(t17)
r17<-c(t17, rep(0,(20-length(t17))))

#**Example 18
t<-optim(c(-0.1, rep(0.001, 18)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 18)), alpha=0.5)
t18<-rv(t$par)
ps18<-ps(t18)
dc18<-DC1(t18)
r18<-c(t18, rep(0,(20-length(t18))))

#**Example 19
t<-optim(c(-0.1, rep(0.001, 19)), psdc.alpha, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 19)), alpha=0.5)
t19<-rv(t$par)
ps19<-ps(t19)
dc19<-DC1(t19)
r19<-c(t19, rep(0,(20-length(t19))))
size<-seq(2,20,1)
MDPs<-rbind(r1,r2,r3,r4,r5,r6,r7,r8,r9,r10,r11,r12,r13,r14,r15,r16,r17,r18,r19)
MDPs
psepDPs<-c(ps1,ps2,ps3,ps4,ps5,ps6,ps7,ps8,ps9,ps10,ps11,ps12,ps13,
ps14,ps15,ps16,ps17,ps18,ps19)
ps1DPs<-round(psepDPs,digits=4)
prsepDPs<-ps1DPs  #This will be used to draw line graph
dcDPs<-c(dc1,dc2,dc3,dc4,dc5,dc6,dc7,dc8,dc9,dc10,dc11,dc12,dc13,dc14,
dc15,dc16,dc17,dc18,dc19)
dc1DPs<-format(dcDPs,digits=4,scientific=TRUE)
detinfoDPs<-1/dcDPs
detinfoDPs
dc1DPs0.5<-dcDPs
detinfoDPs0.5<-detinfoDPs
```

R Codes for Simulation Studies of non-Bayesian Designs

```
optsep3=function(NS,n,alpha,beta0,beta1){
OEsti=matrix(0,nrow=NS,ncol=2)   #original estimates
Bias=matrix(0,nrow=NS,ncol=2)
BiasSq=matrix(0,nrow=NS,ncol=2)
sep=rep(0,times=NS)
CS=rep(0,times=NS)
resmatrix=matrix(0,nrow=NS,ncol=n)
t<-optim(c(-0.1, rep(0.001, (n-1))), psdc.beta, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, (n-1))), alpha, beta0,
    beta1)
beta<-c(beta0,beta1)  #coefficients
intercept=rep(1,n)
z=cbind(intercept,x1)
pred<- z%*%beta   #Multiplication of a matrix with a column vector
p<- 1/(1+exp(-pred))  #calculation prob in logistic regression
for(s in 1:NS){
y=NULL
for(k in 1:n) {
y[k]<- rbinom(1,1,p[k])
              }
resmatrix[s,]=y
CS1=rep(0,times=n)
for(k in 1:n){
if(sum(y[1:k])==n) {CS1[k]=1}
else if(sum(y[1:n])==0) {CS1[k]=1}
else if(sum(y[1:k])==0 & sum(y[(k+1):n])==(n-k)){CS1[k]=1}
else if(sum(y[1:k])==k & sum(y[(k+1):n])==0){CS1[k]=1}
else {CS1[k]=0}  #counting number of complete separation
            }
CS[s]=sum(CS1)
```

```
#Generated data:
w1=cbind(y,x1)
results<- glm(y~x1,family=binomial)
k0=results$coefficients[1]   #estimate of intercept
k1=results$coefficients[2]   #estimate of the coefficient of x1
TP=beta  #True parameters
OEsti1=c(k0,k1)
OEsti[s,]=OEsti1
B=OEsti1-TP #Bias of the estimates
Bias[s,]=B #Bias of the estimates in the ith simulations
BiasSq[s,]=B^2 #Bias square of the estimates in the ith
        }
OEstimates=apply(OEsti,2,mean,na.rm=TRUE)
SEEst=apply(OEsti,2,sd,na.rm=TRUE)  #Standard errors of estimates
BiasEst=apply(Bias, 2, mean,na.rm=TRUE)
RelBias=(BiasEst/TP)*100  #Relative bias in percetage
MSE=apply(BiasSq, 2, mean,na.rm=TRUE)
RMSE=sqrt(MSE)
#Estimates based on median
OMedEstimates=apply(OEsti,2,median,na.rm=TRUE)
BiasMd=OMedEstimates-TP
#Median square error
MdSE=apply(BiasSq,2,median,na.rm=TRUE)
IQRMd=apply(OEsti,2,IQR,na.rm=TRUE)
ComS=mean(CS)*100
#EstMatrix=cbind(OEstimates,SEEst, BiasEst,RelBias,MSE, RMSE)
Fresults=cbind(beta0,beta1,alpha,BiasEst[1],BiasEst[2], SEEst[1],
SEEst[2],MSE[1],MSE[2],ComS)
library(xtable)
FMdresults=cbind(beta0,beta1,alpha,BiasMd[1],BiasMd[2], IQRMd[1],
IQRMd[2],MdSE[1],MdSE[2],ComS)
xmedian=xtable(FMdresults,digit=c(1,2,2,   2,3,3,3,   3,3,3,2))
print(xmedian)
print(ComS)
}
optsep3(NS=10000,n=8,alpha=0.25,beta0=0,beta1=0.5)
optsep3(NS=10000,n=8,alpha=0.50,beta0=0,beta1=0.5)
optsep3(NS=10000,n=8,alpha=0.75,beta0=0,beta1=0.5)
optsep3(NS=10000,n=8,alpha=0.80,beta0=0,beta1=0.5)
optsep3(NS=10000,n=8,alpha=0.90,beta0=0,beta1=0.5)
optsep3(NS=10000,n=8,alpha=1.00,beta0=0,beta1=0.5)
```

R Codes for Computation of Pseudo Bayesian Designs

```
#Optimal Bayesian design
R=5000
meanb0=0
sdb0=0.25
meanb1=1
sdb1=0.25
beta0=rnorm(R,meanb0,sdb0)
beta1=rnorm(R,meanb1,sdb1)
TP=c(meanb0,meanb1)
psdc.SB=function(g, alpha,R){
for(m in 1:R){
s <- NULL; f<- NULL; x<-NULL; d<-NULL; w<-NULL; pcs<- NULL;
n<- length(g)
x[1] <- g[1]
for(k in 1:(n-1)){
x[k+1]<- x[k]+g[k+1]
                }
for(i in 1:n)  {
s[i]<- 1/(1+exp(-(beta0[m] + beta1[m]*(x[i]))))
f[i]<- 1- s[i]  }
#Cumulative product of s
```

```
cps<-cumprod(s)
#Cumulative product of f
cpf<-cumprod(f)
for(k in 1:(n-1)){
pcs[k]<- (cpf[k]*(cps[n]/cps[k]) + cps[k]*(cpf[n]/cpf[k]))
                  }
ps<- cps[n]+sum(pcs)+cpf[n]
x11<-rep(1, n)
x12<-x
X<-cbind(x11,x12)
for(t in 1:n){
            w[t]<-(1/n)
              }
wp<-w*s*f
W<-diag(wp)
Xt<-t(X)
INF<-Xt %*% W %*% X
DINF<-det(INF)
DCO<-(1/DINF)^(alpha/2)  # this 2 is for no. of parameters
PSO<-(ps)^(1-alpha)
#Combined criterion
DCSB[m]<-(PSO)*(DCO)
DCSB1<-sum(DCSB,na.rm=TRUE)*(1/R)
}
cat("The value of DCSB1 criterion\n")
print(DCSB1)
}
t<-optim(c(-0.1, rep(0.001, 7)), psdc.SB, NULL, method="L-BFGS-B",
lower=c(-20,rep(0.0001, 7)), alpha=0.25,R=10000)
t2<-rv(t$par)
BDn8a25=t2
xtable(t(as.matrix(BDn8a25)),digits=4)
```

R Codes for Simulation Studies of Pseudo Bayesian Designs

```
optBayes8=function(x1,NS,n,alpha,beta0,beta1){
CEsti=matrix(0,nrow=NS,ncol=2)
OEsti=matrix(0,nrow=NS,ncol=2)
OBias=matrix(0,nrow=NS,ncol=2)
OBiasSq=matrix(0,nrow=NS,ncol=2)
Bias=matrix(0,nrow=NS,ncol=2)
BiasSq=matrix(0,nrow=NS,ncol=2)
CS=rep(0,times=NS)
resmatrix=matrix(0,nrow=NS,ncol=n)
beta<-c(beta0,beta1)  #coefficients
intercept=rep(1,n)
z=cbind(intercept,x1)
pred<- z%*%beta
p<- 1/(1+exp(-pred))
for(s in 1:NS){
y=NULL
for(k in 1:n){
y[k]<- rbinom(1,1,p[k])
             }
resmatrix[s,]=y
CS1=rep(0,times=n)
for(k in 1:n){
if(sum(y[1:k])==n) {CS1[k]=1}
else if(sum(y[1:n])==0) {CS1[k]=1}
else if(sum(y[1:k])==0 & sum(y[(k+1):n])==(n-k)){CS1[k]=1}
else if(sum(y[1:k])==k & sum(y[(k+1):n])==0){CS1[k]=1}
else {CS1[k]=0}  #counting number of complete separation
             }
CS[s]=sum(CS1, na.rm=TRUE)
w1=cbind(y,x1)
```

```
k0=results$coefficients[1]
k1=results$coefficients[2]
g0=abs(k0)
g1=abs(k1)
Esti=c(0,0)
NOS=0
if (g1>20) {Esti=c(NA,NA) } else {Esti=c(k0,k1)}
CEsti[s,]=Esti   #parameter estimates in sth simulation
B=Esti-TP      #Bias of the estimates
Bias[s,]=B     #Bias of the estimates in the ith simulations
BiasSq[s,]=B^2 #Bias square of the estimates in the ith
               #simulations to be used in MSE calculations
OEsti[s,]=c(k0,k1)
OEstim=OEsti[s,]
OBias[s,]=OEstim-TP
OBiasSq[s,]=(OEstim-TP)^2
}
CEstimates=apply(CEsti,2,mean,na.rm=TRUE)
SEEst=apply(CEsti,2,sd,na.rm=TRUE)  #Standard errors of estimates
BiasEst=apply(Bias, 2, mean,na.rm=TRUE)
RelBias=(BiasEst/TP)*100  #Relative bias in percetage
MSE=apply(BiasSq, 2, mean,na.rm=TRUE)
RMSE=sqrt(MSE)
OEstimates=apply(OEsti,2,mean,na.rm=TRUE)
OSEEst=apply(OEsti,2,sd,na.rm=TRUE)  #Standard errors of estimates
OBiasEst=apply(OBias, 2, mean,na.rm=TRUE)
ORelBias=(OBiasEst/TP)*100  #Relative bias in percetage
OMSE=apply(OBiasSq, 2, mean,na.rm=TRUE)
ORMSE=sqrt(OMSE)
#Estimates based on median
OMedEstimates=apply(OEsti,2,median,na.rm=TRUE)
BiasMd=OMedEstimates-TP
#Median square error
MdSE=apply(OBiasSq,2,median,na.rm=TRUE)
IQRMd=apply(OEsti,2,IQR,na.rm=TRUE)
ComS=mean(CS)*100
EstMatrix=cbind(CEstimates,SEEst, BiasEst,RelBias,MSE, RMSE)
Fresults=cbind(beta0,beta1,alpha,BiasEst[1],BiasEst[2], SEEst[1],
SEEst[2],MSE[1],MSE[2],ComS)
FOresults=cbind(beta0,beta1,alpha,OBiasEst[1],OBiasEst[2], OSEEst[1],
OSEEst[2],OMSE[1],OMSE[2],ComS)
library(xtable)

FMdresults=cbind(beta0,beta1,alpha,BiasMd[1],BiasMd[2], IQRMd[1],
IQRMd[2],MdSE[1],MdSE[2],ComS)
xmedian=xtable(FMdresults,digit=c(1,2,2,   2,3,3,3,   3,3,3,2))
print(xmedian)
print(ComS)
}
optBayes8(x1=BDn8a25,NS=10000,n=8,alpha=0.25,beta0=0,beta1=1)
optBayes8(x1=BDn8a50,NS=10000,n=8,alpha=0.50,beta0=0,beta1=1)
optBayes8(x1=BDn8a75,NS=10000,n=8,alpha=0.75,beta0=0,beta1=1)
optBayes8(x1=BDn8a80,NS=10000,n=8,alpha=0.80,beta0=0,beta1=1)
optBayes8(x1=BDn8a90,NS=10000,n=8,alpha=0.90,beta0=0,beta1=1)
optBayes8(x1=BDn8a100,NS=10000,n=8,alpha=1,beta0=0,beta1=1)
```

References

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, pages 131–153. 122

Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley. 75, 76, 88

Agresti, A. and Yang, M.-C. (1987). An empirical investigation of some effects of sparseness in contingency tables. *Computational Statistics & Data Analysis*, 5(1):9–21. 11, 12

Albert, A. and Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71:1–10. 11, 116, 118, 120, 121, 128

Allison, P. D. (2008). Convergence failures in logistic regression. *SAS Global Forum, Paper 360-2008*, http://www2.sas.com/proceedings/forum2008/360-2008.pdf. 11, 12, 117, 118, 119, 122, 123, 176

Altman, M., Gill, J., and McDonald, M. P. (2004). *Numerical Issues in Statistical Computing for the Social Scientist*, volume 508. John Wiley & Sons. 11, 12

Antony, J. and Capon, N. (1998). Teaching experimental design techniques to industrial engineers. *International Journal of Engineering Education*, 14(5):335–343. 3

Atkinson, A. (2008). DT-optimum designs for model discrimination and parameter estimation. *Journal of Statistical planning and Inference*, 138(1):56–64. 134, 136

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press Oxford. 115, 178

Bailey, R. A. (2008). *Design of Comparative Experiments*. Cambridge University Press. 96

Bayarri, M. J. and Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, pages 58–80. 5, 8

Berger, M. P. F. and Wong, W. K., editors (2005). *Applied Optimal Designs*. Wiley. 114

Besag, J. (2001). Markov chain monte carlo for statistical inference. *Center for Statistics and the Social Sciences*. 18

Biedermann, S., Dette, H., and Pepelyshev, A. (2004). Maximin optimal designs for a compartmental model. In *MODA 7 Advances in Model-Oriented Design and Analysis*, pages 41–49. Springer. 165, 168, 169

Biedermann, S., Dette, H., and Zhu, W. (2007). Compound optimal designs for percentile estimation in dose–response models with restricted design intervals. *Journal of Statistical Planning and Inference*, 137(12):3838–3847. 134, 136

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of Royal Statistical Society, Series A*, 143:383–430. 112

Brooks, S. (1998). Markov chain monte carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1):69–100. 18, 20

Bryson, M. C. and Johnson, M. E. (1981). The incidence of monotone likelihood in the Cox model. *Technometrics*, 23(4):381–383. 10, 11

Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24):4279–4292. 48, 52

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208. 177

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, pages 586–602. 165

Chipman, H. and Hamada, M. (1996). Bayesian analysis of ordered categorical data from industrial experiments. *Technometrics*, 38:1–10. 88, 112, 175

Clarkson, D. B. and Jennrich, R. I. (1991). Computing extended maximum likelihood estimates for linear parameter models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 417–426. 11, 12

Collett, D. and Stepniewska, K. (1999). Some practical issues in binary data analysis. *Statistics in Medicine*, 18(17-18):2209–2221. 88, 175

Collins, L. M., Schafer, J. L., and Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, 6(4):330. 48, 49, 50, 51

Congdon, P. (2007). *Bayesian Statistical Modelling*, volume 704. John Wiley & Sons. 26, 56

Cox, D. (1970). Analysis of binary data, 1970. *Methuen, London*, pages 103–8. 122

Davison, A. C. (1997). *Bootstrap Methods and their Application*, volume 1. Cambridge University Press. 46

Dette, H. (1997). Designing experiments with respect to standardizedoptimality criteria. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):97–110. 168

Dette, H., Haines, L. M., and Imhof, L. A. (2003). Maximin and bayesian optimal designs for regression models. Technical report, Technical Report//Universität Dortmund, SFB 475 Komplexitätsreduktion in Multivariaten Datenstrukturen. 165, 168, 169

Donaldson, J. R. and Schnabel, R. B. (1987). Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics*, 29(1):67–82. 45

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*, volume 57. CRC press. 46

Fienberg, S. E. and Rinaldo, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, 137(11):3430–3445. 9, 10, 11, 118, 119, 120, 121

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80:27–38. 13, 122

Fong, Y., Rue, H., and Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics*, 11:397–412. 9

Galindo-Garre, F., Vermunt, J. K., and Bergsma, W. P. (2004). Bayesian posterior estimation of logit parameters with small samples. *Sociological Methods & Research*, 33(1):88–117. 26, 48, 55, 56, 72

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:515–533. 95, 99, 103, 176

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472. 37

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):pp. 881–889. 112

Gilmour, S. G. and Goos, P. (2009). Analysis of data from nonorthogonal multistratum designs in industrial experiments. *Applied Statistics*, 58:467–484. 6, 8, 102

Gilmour, S. G., Pardo, J. M., Trinca, L. A., Niranjan, K., and Mottram, D. S. (2000). A split-unit response surface design for improving aroma retention in freeze dried coffee. *In Proc. 6th Eur. Conf. Food-industry Statist. Paris: Socit Francaise de Statistique*, pages 18.0–18.9. 8

Gimenez, O., Choquet, R., Lamor, L., Scofield, P., Fletcher, D., Lebreton, J.-D., and Pradel, R. (2005). Efficient profile-likelihood confidence intervals for capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(2):184–196. 45

Gloneck, G., Darroch, J. N., and Speed, T. P. (1988). On the existence of maximum likelihood estimator for hierarchical models log-linear models. *Scandinavian Journal of Statistics*, pages 187–193. 10

Goos, P. and Gilmour, S. G. (2012). A general strategy for analyzing data from split-plot and multistratum experimental designs. *Technometrics*, 54(4):340–354. 4, 7, 8, 74, 80, 81, 82, 83, 85, 88, 89, 90, 91, 98, 118, 175

Haberman, S. J. (1974). *The Analysis of Frequency Data*, volume 194. University of Chicago Press Chicago. 9, 10, 11, 120, 121

Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, pages 815–841. 10

Hamada, M. and Tse, S. (1988). A note on the existence of maximum likelihood estimates in linear regression models using interval-censored data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 293–296. 9

Hamada, M. and Tse, S. (1992). On estimability problems in industrial experiments with censored data. *Statistica Sinica*, 2(38):1–39. 3, 10

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21(16):2409–2419. 10, 11, 12, 13, 143

Hirji, K. F., Tsiatis, A. A., and Mehta, C. R. (1989). Median unbiased estimation for binary data. *The American Statistician*, 43(1):7–11. 13

Imhof, L. A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Annals of Statistics*, pages 561–576. 169

Jacobsen, M. (1989). Existence and unicity of mles in discrete exponential family distributions. *Scandinavian Journal of Statistics*, pages 335–349. 11, 12

Jones, B. and Goos, P. (2007). A candidate-set-free algorithm for generating d-optimal designs. *Applied Statistics*, 56:347–364. 77, 80

Jones, B. and Nachtsheim, C. J. (2009). Split-plot designs: What, why, and how. *Journal of Quality Technology*, 41(4):340–361. 7

Kruschke, J. (2011). *Doing Bayesian Data Analysis: A Tutorial Introduction with R and BUGS*. Academic Press. 20

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., and Jones, D. R. (2005). How vague is vague? a simulation study of the impact of the use of vague prior distributions in mcmc using winbugs. *Statistics in medicine*, 24(15):2401–2428. 47, 55, 69

Lamotte, L. R. (2005). A note on separated data and exact likelihood-ratio p-values in logistic regression. *Journal of Statistical Computation and Simulation*, 75(8):667–672. 11, 13

Lesaffre, E. and Marx, B. D. (1993). Collinearity in generalized linear regression. *Communications in Statistics-Theory and Methods*, 22(7):1933–1952. 12

Letsinger, J. D., Myers, R. H., and Lentner, M. (1996). Response surface methods for bi-randomization structures. *Journal of Quality Technology*, 28(4):381–397. 8, 75

Litière, S., Alonso, A., and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, 27(16):3125–3144. 54, 56

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337. 20

McGree, J., Eccleston, J., and Duffull, S. (2008). Compound optimal design criteria for nonlinear models. *Journal of Biopharmaceutical Statistics*, 18(4):646–661. 134, 136

McGree, J. M., Drovandi, C. C., Thompson, M., Eccleston, J., Duffull, S., Mengersen, K., Pettitt, A. N., and Goggin, T. (2012). Adaptive Bayesian compound designs for dose finding studies. *Journal of Statistical Planning and Inference*, 142(6):1480–1492. 136

McGree, J. M. and Eccleston, J. A. (2008). Probability-based optimal designs. *Australian New Zealand Journal of Statistics*, 50:13–28. 133, 135

Montgomery, D. C. (1999). Experimental design for product and process design and development. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 48(2):159–177. 2

Montgomery, D. C. (2008). *Design and Analysis of Experiments*. John Wiley & Sons. 2, 3

Müller, C. H. (1995). Maximin efficient designs for estimating nonlinear aspects in linear models. *Journal of Statistical Planning and Inference*, 44(1):117–132. 168, 169

Müller, W. G. and Stehlík, M. (2010). Compound optimal spatial designs. *Environmetrics*, 21(3-4):354–364. 134, 136

Nestorov, I. A. (1999). Sensitivity analysis of pharmacokinetic and pharmacodynamic systems: I. a structural approach to sensitivity analysis of physiologically based pharmacokinetic models. *Journal of Pharmacokinetics and Biopharmaceutics*, 27:577–596. 155

Ntzoufras, I. (2009). *Bayesian Modelling Using WinBUGS*. John Wiley and Sons. 20, 21, 42

OHagan, A. (1994). Kendalls Advanced Theory of Statistics 2B, Bayesian Inference, Section 2.51. 51

Palta, M. (2003). *Quantitative Methods in Population Health: Extensions of Ordinary Regression*, volume 497. Wiley. 17

Pettit, L. I. (1986). Diagnostics in Bayesian model choice. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 64:183–190. 104

Pinheiro, J. C. and Bates, D. M. (2000). *Linear Mixed-Effects Models: Basic Concepts and Examples*. Springer. 43

Royston, P. (2007). Profile likelihood for estimation and confidence intervals. *Stata Journal*, 7(3):376–387. 46

Santner, T. J. and Duffy, D. E. (1986). A note on A. Albert and JA Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 73(3):755–758. 11, 12

Schabenberger, O. (2005). Introducing the glimmix procedure for generalized linear mixed models. *SUGI 30 Proceedings*, pages 196–30. 81

Searle, S. R. (2012). *Linear Models*. John Wiley & Sons. 17

Silvapulle, M. J. (1981). On the existence of maximum likelihood estimates for the binomial response models. *Journal of Royal Statistical Society, Series B*, 43:310–13. 11, 12, 116

Silvapulle, M. J. and Burridge, J. (1986). Existence of maximum likelihood estimates in regression models for grouped and ungrouped data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 100–106. 9

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64:583–639. 76

Taguchi, G., Clausing, D., and Watanabe, L. T. (1987). *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs*, volume 1. UNIPUB/Kraus International Publications White Plains, NY. 2

Tan, M. H. and Wu, C. (2013). A Bayesian approach for model selection in fractionated split plot experiments with applications in robust parameter design. *Technometrics*, 55(3):359–372. 112

Tang, L., Song, J., Belin, T. R., and Unützer, J. (2005). A comparison of imputation methods in a longitudinal randomized clinical trial. *Statistics in Medicine*, 24(14):2111–2128. 50

Venzon, D. and Moolgavkar, S. (1988). A method for computing profile-likelihood-based confidence intervals. *Applied Statistics*, pages 87–94. 46

Wedderburn, R. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika*, 63(1):27–32. 120

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48(3-4):233–243. 81

Woods, D., Lewis, S., Eccleston, J., and Russell, K. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics*, 48(2):284–292. 135

Zhang, H., Lu, N., Feng, C., Thurston, S. W., Xia, Y., Zhu, L., and Tu, X. M. (2011). On fitting generalized linear mixed-effects models for binary responses using different statistical packages. *Statistics in Medicine*. 104

Zorn, C. (2005). A solution to separation in binary response models. *Political Analysis*, 13(2):157–170. 11